ICAL ROOM
AL LIBRARY
OF MICH.

MAR 24 1924

Journal of Educational Research

OFFICIAL ORGAN OF THE
EDUCATIONAL RESEARCH ASSOCIATION

B. R. BUCKINGHAM, EDITOR

ASSOCIATE EDITORS

E. J. ASHBAUGH W. W. CHARTERS GEORGE D. STRAYER
GUY M. WHIPPLE S. A. COURTIS LEWIS M. TERMAN
HARRY A. GREENE

VOL. IX MARCH, 1924 No. 3

CONTENTS

	Page
REDUCING THE VARIABILITY IN TEACHERS' MARKS — E. J. Ashbaugh	185
FACTORS DETERMINING SALARY SCHEDULES — B. E. Lewis	199
THE USE OF THE CORRELATION GRAPH WITH HALF-SIGMA CLASS INTERVALS — Gertrude Pand	213
THE INFLUENCE OF LATIN ON THE SPELLING OF ENGLISH WORDS — Warren W. Coxe	233
THE RELIABILITY OF TESTS REQUIRING ALTERNATIVE RESPONSES — Wm. Asker	234
EDITORIALS	241
REVIEWS AND ABSTRACTS	245
NEWS ITEMS AND COMMUNICATIONS	254
EDUCATIONAL RESEARCH ASSOCIATION	265

Entered as second-class matter at the Post Office at Bloomington, Illinois under the act of March 3, 1879.

Journal of Educational Research

Copyright 1923, by
Public School Publishing Co.

Official Organ of the Educational Research Association

Editor-in-Chief, B. R. Buckingham
Director of Bureau of Educational Research, Ohio State University, Columbus, Ohio

Associate Editors

E. J. ASHBAUGH, *Ohio State University, Columbus.*

W. W. CHARTERS, *Carnegie Inst. Tech., Pittsburgh.*

S. A. COURTIS, *City Schools, Detroit, Mich.*

HARRY A. GREENE, *Ex-Officio, Secretary Educational Research Association, State University of Iowa, Iowa City.*

GEORGE D. STRAYER, *Teachers College, Columbia University, New York City.*

LEWIS M. TERMAN, *Leland Stanford Junior University, Stanford University, California*

GUY M. WHIPPLE, *University of Michigan, Ann Arbor, Michigan.*

Contributing Editors

VIRGIL E. DICKSON, *Director, Bureau of Research and Guidance, Public Schools, Oakland and Berkeley, California.*

CHARLES FORDYCE, *Director of Bureau of Educational Measurements, University of Nebraska, Lincoln, Nebr.*

M. E. HAGGERTY, *Dean of the College of Education, University of Minnesota, Minneapolis.*

V. A. C. HENMON, *Director, School of Education, University of Wisconsin, Madison.*

ARTHUR W. KALLOM, *Assistant Director, Department of Educational Investigation and Measurement, Public Schools, Boston, Mass.*

F. J. KELLY, *Dean of University Administration, University of Kansas, Lawrence, Kans.*

ROBT. H. LANE, *Assistant Superintendent of Schools Los Angeles, Cal.*

WILLIAM A. McCALL, *Assistant Professor of Education, Teachers' College, Columbia University.*

GEORGE W. MELCHER, *Director, Bureau of Research and Efficiency, Public Schools, Kansas City, Missouri.*

EUGENE A. NIFENECKER, *Assistant Director, Bureau of Reference, Research and Statistics, Public Schools, New York City.*

JOSEPH P. O'HERN, *Assistant Superintendent of Schools and Director of Efficiency Bureau, Public Schools, Rochester, New York.*

LEON O. SMITH, *Assistant Superintendent (in charge of Educational Research) Public Schools, Omaha, Nebraska.*

W. W. THEISEN, *Assistant Superintendent of Schools Milwaukee, Wis.*

CLIFFORD WOODY, *Director Bureau of Tests and Measurements, University of Michigan, Ann Arbor, Mich.*

The JOURNAL OF EDUCATIONAL RESEARCH is published at Bloomington, Illinois, monthly, except July and August by the Public School Publishing Co. The subscription price is $4.00 $4.20 in Canada or Mexico; and $4.30 in other foreign countries.

The editorial office is at Columbus, Ohio. Address all communications on editorial matters to B. R. Buckingham, Director of Bureau of Educational Research, The Ohio State University, Columbus, Ohio.

Subscriptions and all communications about reprints or other business matters should be addressed to the Public School Publishing Company at Bloomington, Illinois.

JOURNAL *of* EDUCATIONAL RESEARCH

Volume IX MARCH, 1924 Number 3

REDUCING THE VARIABILITY IN TEACHERS' MARKS

E. J. ASHBAUGH
Ohio State University

For a number of years a discussion of teachers' marks has had a large place in current literature. A complete bibliography covering the last decade would include the names of more than fifty separate articles. These articles have confined themselves almost entirely to two phases of the subject: first, a demonstration of the variability or unreliability of teachers' marks; and second, the setting forth of some "new" scheme of marking which has appealed to the writer as an improvement.

In the former group may be found the writings of Starch, Rugg, Kelley, and others. They bring forcibly to the attention of superintendents and teachers the fact that high grades given by one teacher or in one school often denote no higher accomplishment than low grades given by another teacher or in another school. In the second group of articles, the writers have devoted themselves largely to arguments for a letter system of five or more points rather than a percentage gradation, and to the idea of variation in credit for differing qualities of work. The evidence seems to show that when teachers grade by letter instead of by a percent system, or by units of not less than five percent instead of by units of one percent, the variability is somewhat reduced.

Steele[1] and Starch[2] attempted to give some guidance in the

[1] Steele, A. G. "Training teachers to grade." *Pedagogical Seminary*, 18:523 ff., 1911.

[2] Starch, Daniel. "Can variability of marks be reduced?" *School and Society*, 2:242-43, 1915.

reduction of variability through having teachers score identical papers, and then examine the resulting variable scores. An occasional article has appeared in which the writer has advocated accurate definition of the grades to be given. Bailey[3] and a bulletin[4] from the Department of Research, Martins Ferry, Ohio, illustrate this type. In both these cases, however, the thought in the mind of the writer has been the marking of students rather than the marking of papers.

With a somewhat different thought in mind from that which other writers have expressed, the following data are presented. They suggest another way in which the variability of teachers' marks may be reduced.

The group under consideration was a class of fifty-five seniors and graduate students in the Department of School Administration at Ohio State University. All the graduate students in this class had had teaching experience, and the seniors were definitely preparing to teach. In order to impress them with the fact of variability in teachers' grading, the following experiment was conducted. Ten problems from Stone's Reasoning Test in Arithmetic were first given to these students as a test, in order to familiarize them with the problems and give them a basis for marking papers. The problems are presented herewith in order that the reader may make easy reference to them.

Stone's Reasoning Problems

1. If you buy 2 tablets at 7 cents each and a book for 65 cents, how much change should you receive from a two-dollar bill?

2. John sold 4 *Saturday Evening Posts* at 5 cents each. He kept ½ of the money and with the other ½ he bought Sunday papers at 2 cents each. How many did he buy?

3. If James had 4 times as much money as George, he would have $16. How much money has George?

4. How many pencils can you buy for 50 cents at the rate of 2 for 5 cents?

5. The uniforms for a baseball nine cost $2.50 each. The shoes cost $2 a pair. What was the total cost of uniforms and shoes for the nine?

[3] Bailey, W. A. "The administration of quantitative and qualitative credit for high-school work." *School Review*, 25:305-21, May, 1917.

[4] *The marking system.* Mimeographed bulletin of Department of Research, Martins Ferry, Ohio, January 18, 1923.

FIGURE 1

Type Paper on Stone Reasoning Test - Arithmetic

(1)

$$\begin{array}{r} 7 \\ 2 \\ \hline 14 \end{array} \qquad \begin{array}{r} 14 \\ 65 \\ \hline 89 \end{array} \qquad \begin{array}{r} 200 \\ 89 \\ \hline \$1.21 \end{array}$$

(2) $5¢ \times 4 = 20¢$, $20¢ \div 2 = 10¢$ $\qquad$ $10¢ \div 2¢ = 5$ papers

(3) $\$16 \div 4$

(4) 2 pencils = 5¢
1 " = $\frac{1}{2}$ of 5¢ = $2\frac{1}{2}$¢
50¢ ÷ $2\frac{1}{2}$¢ = 20 pencils

(5)

$$\begin{array}{r} 2.50 \\ 18.00 \\ \hline \$20.50 \end{array} \qquad \begin{array}{r} \$2 \\ 9 \\ \hline \$18 \end{array}$$

(6) $2200 \div 2 = 1100$, $1100 \div 4 = 275$, $275 \div 8 = 34$ 3/8 $\qquad$ 3/8

(7) $\frac{3 \text{ tons} \times 1.05}{3.5 \text{ tons}} = 30.50$

(8) $\$1.00 \div 5 = 20$ $\qquad$ $1.20 \div 5 = 24$ $\qquad$ $24 - 20 = 4$ magazines

(9) 1/8 + 3/8 = 4/8

4/8 = 80 $\qquad$ 1/8 = 20 $\qquad$ 8/8 = 1.60

(10) 28/70 x 2.10 = 84¢

210 − 84 = 126¢

6. In the schools of a certain city there are 2,200 pupils; ½ are in the primary grades, ¼ in the grammar grades, ⅛ in the high school and the rest in the night school. How many pupils are there in the night school?

7. If 3½ tons of coal cost $21, what will 5½ tons cost?

8. A news dealer bought some magazines for $1. He sold them for $1.20, gaining 5 cents on each magazine. How many magazines were there?

9. A girl spent ⅛ of her money for car fare, and three times as much for clothes. Half of what she had left was 80 cents. How much money did she have at first?

10. Two girls receive $2.10 for making button-holes. One makes 42, the other 28. How shall they divide the money?

TABLE I. TOTAL SCORES BY EACH STUDENT AT SUCCESSIVE TRIALS

Student's Number	First Trial	Second Trial	Third Trial	Student's Number	First Trial	Second Trial	Third Trial
1	80	60	53	26	54	76	70
2	73	57	55	27	54	54	54
3	70	80	60	28	53	55	63
4	70	69	65	29	52	45	50
5	68	65	73	30	52	40	40
6	67	65	58	31	51	58	60
7	65	67	65	32	51	60	50
8	64	73	66	33	50	60	62
9	63	52	40	34	50	58	74
10	63	40	38	35	48	64	62
11	62	50	65	36	48	44	57
12	61	74	60	37	48	44	35
13	61	70	70	38	47	58	63
14	61	55	70	39	47	45	50
15	61	53	67	40	46	60	56
16	61	49	35	41	46	45	55
17	60	70	60	42	45	57	55
18	60	56	44	43	43	40	60
19	60	55	55	44	42	47	54
20	60	53	48	45	40	48	58
21	58	50	60	46	40	40	50
22	58	50	58	47	40	40	49
23	55	55	53	48	36	40	40
24	55	44	48	49	29	30	42
25	55	30	40				

Each student, after taking this test, received a mimeographed copy of an answer paper. This paper is reproduced in Figure 1, and will be referred to hereafter as the Type Paper. It was taken from the work of a class of seventh-grade pupils. The instruction to the students was to score this paper on the basis of

ten points for each problem, and with the assumption that it had been written by a seventh-grade pupil in the scorer's own school.

When the scoring has been completed, each member of the class wrote his name on the sheet he had scored, and the papers were collected. No comment was made on the scoring, and the class continued as if this work had not been done. Four weeks later, the class was given new copies of the Type Paper with the same instructions as before, namely that it be scored on the basis of ten points for each problem and with the assumption that it had been written by a seventh-grade pupil in the scorer's own school. Again the papers were signed and passed in. Four weeks later a third scoring of the Type Paper was made by the same students and under the same instructions. Due to absences, only 49 students handed in all three scorings. The total number of points allowed to the Type Paper by each of these 49 students at each trial is given in Table I. The table is read thus: In scoring the Type Paper, student number one gave a total score of 80 on the first trial, 60 on the second, and 53 on the third. A summary of these scores is given in Table II.

TABLE II. DISTRIBUTION OF TOTAL SCORES GIVEN ON TYPE PAPER

Scores	First Trial	Second Trial	Third Trial
80–84	1	1	..
75–79	..	1	..
70–74	3	4	5
65–69	3	4	4
60–64	13	5	11
55–59	5	10	9
50–54	9	6	8
45–49	8	7	3
40–44	5	9	6
35–39	1	0	3
30–34	..	2	..
25–29	1	..	..
Median	56.0	55.3	57.5

Not until after the third scoring had taken place was there any discussion regarding the basis for grading. It will be noted that under these circumstances the median remained practically constant but that the range of distribution tended to narrow.

Most interesting, however, from the standpoint of the consistency of the individual scorers are the facts of Table III. This table shows the direction but not the amount of the variation from the first to the second and from the second to the third scoring. The table reads: Of the students whose second score was less than their first, 9 gave a third score which was less than the second; 2, a third score which was equal to the second; and 14, a third score which was greater than the second.

TABLE III. DIRECTION OF CHANGE IN VALUE OF TOTAL SCORE

Second Score Compared with First	Third Score Compared with Second			Total
	Third Less Than Second	Third Equal To Second	Third Greater Than Second	
Second less than first....	9	2	14	25
Second equal to first....	1	1	2	4
Second greater than first	10	2	8	20
Total.....	20	5	24	49

It will be noted from this table that but one student of the 49 gave the same total score on all three occasions and that only 7 others gave the same total score on two *successive* occasions. If to the latter are added the three students, shown in Table I (Numbers 7, 17, and 22), who gave the same score on the first and last trials, a total of eleven students is obtained who agreed with themselves to the extent of giving the same score twice. Twenty-four, or almost exactly half, reversed the direction of their variation. In other words, they either dropped at the second scoring and tended to return to the original level or they rose on the second and tended to drop back to the original level on the third. On the other hand, there were 17 students who, on the third trial, continued to vary in the same direction as that which they showed on the second trial. In other words, if their second score was less than the first, their third was still less; and if they increased their score on the second trial, they still further increased it on the third trial.

The average difference between the first and second scores was 8.1 points, with a range of 0 to 25. The average difference between the second and third scores was 7.3. It is inter-

esting to note that one student (Number 25, Table I) gave a score of 55 on the first trial and dropped to 30 on the second, while another (Number 26) gave a score of 54 on the first trial and raised the score to 76 on the second. If we had but the first trial to rely upon, we should say that these students agreed very closely. The second trial, however, finds them 46 points apart. In both cases on the third scoring the tendency was to return toward an intermediate point, thus reducing the disagreement but still leaving it very wide. Number 25 raised from 30 to 40 and Number 26 dropped from 76 to 70. Results of a different type were shown by student Number 10. This student, after giving 63 on the first trial, dropped sharply to 40 on the second and reduced that score two more points on the third trial.

TABLE IV. INCONSISTENCY OF GRADING ON INDIVIDUAL PROBLEMS

Number of Problem	Same Score on All Three Trials	Same Score on Two Trials	Different Score on Each Trial
1.....	11	25	13
2.....	43	4	2
3.....	34	13	2
4.....	35	11	3
5.....	20	23	6
6.....	36	9	4
7.....	23	20	6
8.....	15	27	7
9.....	15	26	8
10.....	26	17	6

The coefficients of correlation (Spearman footrule) between the scores on the different trials were as follows: first and second, $r = 0.50$; first and third, $r = 0.28$; second and third, $r = 0.64$. As a commentary on the lack of self-agreement on the part of those who rate test papers in the usual way, these figures are rather illuminating. All these values of r may certainly be considered too small to be satisfactory.

The inconsistency of grading becomes even more noticeable when attention is directed to the score on individual problems. Table IV presents a summary of this situation. For example in the case of problem one, 11 students gave the same score

on all three trials, 25 gave the same score on two trials, and 13 gave a different score on each trial. It will be noted that the greatest inconsistency was in scoring problem one and the least on problem two.

Table V presents the data in still more detail.

TABLE V. INDIVIDUAL VALUES GIVEN TO EACH PROBLEM BY 49 STUDENTS

	Score	Problem Number									
		1	2	3	4	5	6	7	8	9	10
First Trial	10	3	46	38	41	2		6	21	16	29
	9	1	1	3	3			1			3
	8	6	1	5	2		1	1	2		2
	7	4		1		1	1	1			1
	6	5				2		1			1
	5	9		2	2	4	1	3	3	7	5
	4	2	1		1	2				4	
	3	5				4		2			
	2					2	1	1	2	2	
	1						1				
	0	14				32	44	33	21	20	8
Second Trial	10	1	48	41	45	1	2	3	17	13	35
	9			3							1
	8	2	1	3	1			3	1	1	2
	7	5		1						1	
	6	2					1	2		1	1
	5	15		1	3	6	1	3	10	10	2
	4	5				2		2		2	
	3	3				5		1	1		
	2					2	3	1		3	
	1										
	0	16				33	42	34	20	18	8
Third Trial	10		44	42	44			4	20	10	37
	9		2	5					1	1	
	8	3	2	2	2	1	1	1	1		4
	7	6	1		1		1	1	2		1
	6	3				2		1	2	1	
	5	11			1	7		9	8	12	1
	4	3				6	1			2	
	3	4				2			1	1	1
	2	2				1	4	1		2	
	1									1	
	0	17			1	30	42	32	14	19	5

The table is to be read as follows: On the first trial, three students gave to the solution of problem 1 a score of 10, or perfect; one student gave it a score of 9; six gave it 8; four gave it 7; and so on, with fourteen giving it zero, or stating

that in their judgment the solution was entirely worthless. The reading for other problems is similar. The meaning of the variation in rating may be brought out with reference to problem 2—the problem in which the greatest consistency was shown. On this problem, 46 students gave a perfect score on the first trial. On the second trial, however, 48 considered the solution of problem 2 perfect, the students who had previously rated it 8 and 9, raising their scores to 10. The student who had previously rated it 4 now doubled his rating. On the third scoring only 44 students considered the solution of problem 2 perfect. Three of those who had been satisfied with its perfection on both the first and the second scorings now found some imperfection; while two students were constant in their inconsistency and assigned a still different value on the third scoring from either they had assigned before.

With further reference to Table V, it will be noted that on the first scoring, the solution of every problem except the sixth was regarded as a perfect solution by at least two students; and that, on the other hand, six of these nine problems found a considerable group who considered these same solutions entirely worthless. On the second scoring *every* solution was regarded as perfect by one or more students; yet seven of these solutions were considered worthless by rather large numbers of the students. The solution of problem 8 was given the highest possible credit by 17 of the students and no credit at all by 20 other students. Similarly the solution of problem 9 met every requirement, according to 13 of the students; but it had no merit at all according to 18 others. On the third scoring, as has already been pointed out, the range was not quite so great. On this occasion problems 1, 5, and 6 failed to get a perfect score from any individual, but number 4 fell from its previous high standing and now received at least one vote for zero quality.

It is probably needless to state that when these data were presented to the class, the students were completely convinced of the variability of teachers' marks. However, the experi-

ment was not considered complete at this point. The writer wished to find what could be done by way of reducing this variability. He therefore followed this revelation of existing conditions with a class discussion concerning the bases for scoring papers. An examination of the Type Paper revealed certain interesting facts. If the paper were scored solely with reference to whether the final answer was right or wrong, the paper should receive a score of 60, since problems 5, 6, 7, and 9 alone do not reveal correct answers. If, however, the scoring is on the basis of whether or not the child understood the correct principle involved, the score would be slightly different.

It is evident that on problem 5 the child understood perfectly that he was to find the cost of the shoes and add this to the cost of the uniforms in order to get the total. However, he apparently failed to note the word "each" in the first sentence and therefore found the cost of the shoes for the entire nine and added that to the cost of a single uniform. Since he has thought correctly in at least two of the three processes involved, he might well be accredited with a score of 7.

In problem 6 the child has failed to understand that each of the fractional parts are parts of the total enrollment. Moreover, one cannot be sure whether the child understood anything as to the way to proceed or whether he was simply juggling with the figures. In view of the fact that from three to five steps must be taken in the solution of this problem, dependent upon the method employed, a maximum score of two or three points might be allowed by giving the child the benefit of doubt and assuming that he understood at least the first step in the solution.

On problem 9 the child has recognized that he must bring together the amount spent for car fare and for clothes, but he has failed to get the significance of the second sentence. This is that "*half* of what she had left was 80 cents." Since the first and last of the necessary steps in the solution of the problem are correct, while the middle step has been omitted, it would seem that a grade of 5 or 7 might be allowed.

A question may be raised on number 8. The principles involved are selling price minus cost equals gain, and total gain divided by gain on each article equals number of articles. While $\frac{1.20}{5} - \frac{1.00}{5}$ gives the same answer as $\frac{1.20 - 1.00}{5}$ the separation of the items would with most teachers, we believe, justify the conclusion that the child had not thought the principles through correctly.

Therefore, if we score this paper on the basis of principles involved or correct thinking, a score of 74 to 77 will result. This may be compared with the rating of 60 which was suggested above on the basis of the correct answer.

Again, if we score the paper so as to give no value to a solution which involves any type of error,—with the exception of analysis or accurate mathematical statement—we shall find that only solutions two, three, four, and ten pass muster, and that a rating of 40 results.

To the one who wished to check from the standpoint of analysis or exact accuracy of mathematical statement, even these fail to measure up. In the solution of problem two, 4×5 cents does not give 20, but 20 cents; and 10 cents divided by 2 gives 5 cents, and not 5 papers. Likewise, in problem 3, \$16 divided by 4 gives \$4, and the dollar sign has been omitted. In number 4, since the answer desired is pencils and not cents the first two statements are stated in reversed order. Also, 50 cents divided by 2½ cents does not equal 20 pencils. In number 10, $\frac{28}{70} \times 2.10 = 84$, not 84 cents, and $210 - 84 = 126$, an abstract number.

After the solutions of the problems had thus been carefully analyzed, the unanimity of scoring was practically perfect. The students agreed that if we decide that papers should be marked wholly on the basis of correct numerical answer, disregarding dollar signs, etc., this paper should be scored 60. A large number of the class, however, were of the opinion that that was not the proper basis for grading. On the other hand, a great

many were positive that a child should not be given a zero score on number 2 because he had designated his answer as papers in a quotient where the corresponding dividend was cents. They were certain that in number 3 zero should not be given for the omission of the dollar sign, etc.

As a result of our discussion it was agreed that the following scheme for rating each problem would be fair: 4 for correct computations; 5 for correct principle; and 1 for no variation from correct mathematical statement. On the basis of this agreement, the papers were handed again to the students for rating, with the result given in Table VI.

TABLE VI. DISTRIBUTION OF TOTAL SCORES AFTER DISCUSSION OF BASIS FOR SCORING

Scores	Frequency
85–89....	2
81–84....	3
75–79....	6
70–74....	35
65–69....	3
Median	74

Three things should be pointed out in connection with this table:

1. The total range of points is now only 18 (68-86) as compared with 51 on the first scoring, 50 on the second, and 39 on the third. In other words, the range of variability has been reduced to less than half of the best previous scoring.

2. The number of cases varying widely from the median has been reduced very materially. The number of cases with a greater range than plus or minus five from the median is now 6, as compared with 32 to 35 in the earlier scorings.

3. The median score has been very materially raised. This is due to the fact that when the scoring was not done on the basis of a certain value for correct reasoning, another for accuracy of operation, and still another for accuracy of mathematical statement, the tendency was to discount very heavily

for lack in either one of these three characteristics. There was also a tendency on the part of a few to give no value to a problem regardless of the accuracy of the numerical work or the reasoning process when there was a lack of completeness in analysis or failure to follow exact mathematical statement.

In the correction of arithmetic papers the writer has no desire to attempt to justify in detail the division of 5, 4, 1, given above. He is satisfied, for the purpose of this article, to stand on the fact that after considerable discussion, the members of the group felt that this division was fair. The point to this part of the argument is that after a careful, detailed discussion of the items involved in the marking of test papers, which in themselves are not wholly objective, an agreement among those who are to do the marking upon the values to be given to certain phases of the work will result in reduced variability. It will be seen that this procedure is easily adapted to work among the teachers of a single department where a common examination may be given by the entire group.

While not having the detailed data present to prove the point, the writer knows from experience that the same procedure will bring similar results when applied to the entire teaching force of a high school or of the various grades of a school system. As high-school principal and superintendent in small cities for a number of years, he followed this procedure in defining definitely and accurately the characteristics which should be considered in the scoring of examination papers. This was largely before the days of standardized tests; but the results in preventing high grades by one teacher and low grades by another with consequent feeling on the part of students that certain teachers were to be avoided while others were "easy," made for better morale and scholastic work. To the extent that few items are considered in the final scoring and these items are clearly defined, will the teachers' marks closely approximate the students' accomplishment in an examination.

The discussion in this article has been concerned with the marking of examination papers and not with the grading of

students for promotion. It seems highly desirable that these two ideas should be kept separate. Probably when the question of promotion is involved, personality, attitude, persistence, industry, etc., should be given consideration; but students have a right to expect that their marks on examination papers shall portray accomplishment on the task assigned.

The use of five-point intervals instead of single percents and of letter groups instead of numerical values, the publication of the distribution of teachers' marks in open bulletins, the insistence that each teacher tabulate and plot her distribution of marks before submitting them to the office—all these have been recommended in previous articles and they are of value. An agreement on the part of the teachers to a clear-cut definition of what is being marked and of the relative value to be assigned to the various elements will do even more in reducing the variability of teachers' ratings and in making them a valuable adjunct in school administration.

FACTORS DETERMINING SALARY SCHEDULES

E. E. Lewis
Superintendent of Schools, Flint, Michigan

Perhaps the most important development concerning the wages of teachers during the past fifty years has been the idea of a salary schedule. Before the Civil War in a few large cities, notably Boston, Philadelphia, and New York, lists were made of the salaries paid to teachers and these lists were used by board members and officers as a basis for determining advances in salaries, replacement salaries, and salaries of beginners. Such lists were not schedules as we now think of the term "schedule." But they were the forerunners of the schedules that developed later.

Previous to the development of the idea of a salary schedule it was the practice, on the part of teachers, to get what they could get or take what they were given, and, on the part of boards of education, to pay what they had to pay or what they could afford to pay. These practices still prevail in hundreds of American communities where the salary schedule idea has never been broached. Each teacher gets what he or she can get. Personal, social, political, charitable, and even religious pulls are exercised. Frequently, the schools get into politics and local bosses dictate not only the selection of the personnel, but also the remuneration. Obviously, under such a plan or rather lack of plan, many injustices are done.

Advantages of a Salary Schedule

The following advantages may be offered in favor of the salary schedule.

It is impersonal.—It tends to fix an impersonal basis for the remuneration of teachers, thus lessening the influence of personal, local, social, charitable, religious, political, or other dangerous influences that operate in public service. This is probably the chief administrative argument. Without a schedule, officers and boards are frequently unjustly subjected to tre-

mendous pressures in the interest of some individual. With a schedule, it is possible to convince all but the most obdurate that justice and fair dealing prevail. Unfortunately, however, unless merit is recognized, a schedule may level all teachers to one basis. At first thought this seems democratic; but it is likely to result in a deadening influence since it reduces competition.

It deals justly with new teachers.—A salary schedule makes it possible to deal justly and fairly with new teachers. New teachers ought not to come into a school system at a higher salary than those they replace unless their experience, training, and ability merit the advance. Every year the superintendent has to fill vacancies. The market for teachers, like all markets, is a shifting one. There seems to be no way to stabilize it. With a fixed salary schedule, the superintendent knows exactly what he can offer for a given amount of training, experience, and merit. Teachers long in service are therefore assured that young and inexperienced teachers will not be allowed to compete with them unfairly.

It deals justly with home teachers.—A salary schedule protects the home teacher against unjust discrimination when she is placed in competition with teachers from outside of the city or locality. It is a fact that many home teachers are willing to work for less in their home town, and that many boards want to employ as many home teachers as possible in order to save taxes. A schedule based upon the cost of living for home teachers will be lower than a schedule based upon the cost of living for imported teachers. Now, the schools are more than local enterprises. The home girl should be able to earn as much at home as away from home. Most schedules recognize this fact and base their minimum upon the living costs of an outside teacher. This protects the home teacher who constitutes, on the average, from fifty to seventy-five percent of the teaching staff in most localities. It may be that too much has been said about inbreeding. It is impossible to find a scientific study of its effect.

It makes a career possible.—A salary schedule helps to make teaching a profession rather than an occupation, a trade,

or a job. The prospective teacher knows beforehand about what she can expect at each stage in her career. It stimulates the recruiting of the best young people; it tends to hold them in the profession and to offset the attractions of other occupations. It is a decided aid in the vocational guidance of teachers.

It secures better preparation.—It stimulates the prospective teacher to prepare more deliberately for teaching and to continue her preparation after entering the profession. Especially is this true where regard for training is an essential consideration in the schedule. Thus, it reduces the future percent of mechanical and routinized job-holders. It is also an important factor in remaking old teachers by rewarding them for additional preparation.

It increases tenure.—Not only is the tenure in the local system increased by a proper salary schedule, but the tenure in the profession, regardless of place, is increased. The almost universal adoption of salary schedules in cities means that a city teacher is guaranteed something of a career regardless of where she teaches. County and state salary schedules would give this same guarantee to teachers in other than city schools. Thus far no progress has been made in the establishment of such schedules. Teachers' associations should set up schedules for the guidance of non-city schools.

It stimulates teachers.—A schedule tends not only to relieve the teacher from constant fear for her economic security, but also to stimulate her to reach the highest possible professional achievement. A guaranteed economic security is a powerful stimulus to the right-minded. Economic insecurity threatens the professional worker and often drives her to the acceptance of low standards of personal culture.

It improves the social status of teachers.—The principles of a just salary schedule are not only in accordance with the principles and practices now prevailing in commercial and industrial activities and in other professions, but they are in accord with universally accepted ideas of justice. The adoption of a schedule

based upon such principles elevates the economic, cultural, and social status of teachers.

It helps fix the yearly budget.—When a salary schedule is in operation it is possible to determine more readily the item in the budget covering the cost of instruction than would otherwise be the case. In fact, the amount can be determined years ahead.

Basic Principles of Salary Schedules

An examination of existing salary schedules reveals no two exactly alike. There seem to be some problems peculiar to each locality. Whether these problems are real or imaginary it is difficult to say. Probably, much of the difference arises from the fact that each school system is a law unto itself and attempts to solve its problems in its own way. Many schedules are formulated by those who have some hobby or bias that they believe should be recognized. Trial-error-and-success methods have been used. The schedule of the nearby cities or the cities of a given class have been examined and from those a schedule for a given locality has been formulated with special reference to the local difficulties. The writer has found from his own experience that many unforeseen situations arise that must be temporized with in the best manner possible. No doubt, this also accounts in part for the great diversity of ideas to be found in salary schedules. There are, however, certain basic principles that are common to all situations and which should not be ignored. A schedule should not be copied simply because it has been successful elsewhere. Sound principles should first be established and a schedule built upon them. The following eight principles are, in the judgment of the writer, fundamental and should, so far as possible, be recognized in every schedule. Some of these principles have been stated by Evenden. The writer has merely modified and abbreviated Evenden's statements.

1. Every teacher is entitled to a minimum salary which will provide a living wage for twelve months and a standard of living which will draw people of refinement and ability into the profession.

2. Other factors being the same, the teacher in one grade should receive as much salary as the teacher in any other grade or school division.
3. Salary increases and attainable maxima should be so arranged that they (a) offer a career in teaching, (b) induce the best young men and women from the high school to prepare for teaching, and (c) *secure constant improvement during the time of teaching.*
4. The more and the better the academic and professional preparation of a teacher, other factors being equal, the more salary she should receive.
5. The more *successful* experience a teacher has had, other factors being equal, the more salary she should receive.
6. Every *successful* teacher should find it possible to pass from a mere "living" wage to a "saving" wage (economic independence) and from that to a "cultural" wage. The second of these, the saving wage, should provide the teacher with a salary adequate to meet her necessary expenses and those of her dependents with enough margin to provide for necessary professional advancement and above that a margin for saving and investment. The third, or "cultural" wage, should be enough to provide for economic independence and still allow for travel, additional study, and a reasonable amount of opportunity for the enjoyment of the best music, literature, and art. The teacher may thus be in a position to transmit the best social inheritance of the race.
7. There should be enough flexibility in the salary schedule to provide extra pay for teachers of extra ability. In other words, merit should be recognized, other factors being equal.
8. Between the sexes there should be, wherever possible, equal pay for equal merit, equal training, equal experience, and equal work.

Economic Problems that Arise in Formulating a Salary Schedule

In formulating a salary schedule many economic and educational problems arise. Among the economic problems the following are included.

How are the teachers now paid?—That is, how much is now paid to each teacher in the system and on what basis? The superintendent should have a record of (a) the experience, (b) the training, and (c) the salary of each teacher in the system under his supervision. The schedule which he proposes to recommend must be built in part upon these data. He will find it practically impossible to lower the salary of any teacher and, in most cases, *highly* undesirable. On the other hand, some teachers will be ahead of practically any schedule that resources permit. They should be asked to stand still until the others catch up. They should not be asked to take a lower salary, since a lower salary might cause them to leave the city; nor is it reasonable to allow them to go ahead until the others are cared for. This therefore, means careful personnel work on the part of the supervisory staff.

How much money is available?—The second economic problem is more difficult. How much money is available this year and how much will be available next year and so on during the life of the schedule? Surely a year-to-year schedule is undesirable. A schedule should be built for the future as well as for the current year. This means that the superintendent must carefully compute the cost of the schedule over a period of, say, ten years. He must remember that the schedule has to be sold not only to the Board of Education, but also to the public and not for a day alone, but for the future. Otherwise his work falls and not only is he discredited, but also the idea of a salary schedule as well.

What should be the minimum salary?—The theoretical answer is: at least enough to provide a living wage for twelve months and a standard of living that will draw people of refinement and ability to teaching. But, practically, how much is this? Who can answer? Certainly, big salaries should not be paid to beginners. The beginner is an apprentice. He should be expected to begin low in teaching just as in other lines. Too often the mistake is made of raising the minimum without much reference to the maximum. The minimum should be at least equal

to the minimum in other occupations requiring equal ability and equal preliminary training. If it is slightly above the minimum of other occupations, it will tend to attract superior ability into teaching. However, since economic conditions change from time to time only a rough average can be applied for a given period of years. If the schedule is properly arranged it will be possible to apply a flat percent increase or decrease as economic conditions change. This can be done without changing a single principle of the schedule.

How fast should one be able to advance?—The rate of advancement is another difficult problem. If a teacher can pass from a beginner's salary to the maximum salary in from five to ten years and must then stand still the rest of her professional career or change her location, there is little incentive for her continued efficiency, especially if she is a home teacher. Some provision should be made for her to advance almost indefinitely.

What should be the relation of the maximum to the minimum?—In most salary schedules the maximum is only one and one-half times the minimum. If the minimum is $1000, the maximum is around $1500. In many schedules the maximum is even less than this. Why? Is it lack of money or lack of foresight in planning the schedule? Probably both reasons account for it, but whatever the cause the closeness of the maximum to the minimum in most schedules is unfortunate. Relatively speaking, under such a schedule, beginners are overpaid and teachers of long service are underpaid. It would be better to have a lower minimum, provided, of course, it was not too low to attract the best beginners. Broadly speaking, the writer is of the opinion that the maximum should be so arranged that it cannot be reached too soon, but that when it is reached it will amount to at least twice the minimum and preferably two and one-half or three times the minimum. With a high maximum the required standard of training and merit should be equally high.

What should be the relation of the maximum to the cost of living?—As indicated above, the teacher of high attainments should be able to advance from the living wage of a beginner to a saving wage and finally, to a cultural wage. Just how to define any of these terms, is, of course, difficult. A living wage for one may be a saving wage and possibly a cultural wage for another. Average estimates must be made. Roughly, we may say that a reasonable saving wage is two, two and one-half, or possibly three times a living wage.

What should be the relation of the schedule to supply and demand?—Undoubtedly, there are always good teachers willing to work cheaper than those employed. The teaching market is like all labor markets in this respect; there is a "bear" and a "bull" market as economic conditions change. A schedule helps to stabilize the market and renders the operation of "bear" and "bull" movements less effective. Some, who cannot or will not "see" the schedule idea, want the hiring of teachers to proceed solely upon the economic principle of supply and demand. Teaching is not primarily an economic undertaking; it is or should be a profession. A profession puts the public interest before the self-interest. It seems reasonable, therefore, to protect teachers to some extent, against the ruthless operation of the principle of supply and demand; but too much protection may be equally dangerous to the profession. Competition is the life of trade.

Educational Problems that Arise in Formulating a Salary Schedule

In addition to the economic problems discussed above, certain educational problems arise that must not be overlooked. We shall discuss some of them.

How many schedules are necessary?—In New York City, teachers and other school officers are paid according to no fewer than eighty-five different salary schedules. This is too many. It is doubtful if there are eighty-five different grades of service in public education. In smaller cities the question arises: Should

there be a separate schedule for kindergarten, primary, intermediate, junior high-school, and senior high-school teachers or, should these distinctions be ignored and a unit salary schedule adopted? The writer feels very strongly the desirability of a unit schedule. The distinction in salary should not be in terms of the grade or group of grades taught. Three things should be considered, namely: (a) the amount of experience; (b) the amount of training; and (c) the merit of each teacher. The particular grade or subject is not so important. If less training is required for a given position in comparison with another, less money should be paid, other factors being equal.

In the older type of salary schedule for example, that of New York City—advances in salary were based upon (a) length of service, and (b) the grade taught. No provision was made for variations in the amount and kind of training or merit. Too much emphasis was placed upon experience and grade taught. Up to 1915, a total of 547 cities reported the adoption of salary schedules, mostly of this older type. By 1922, this number had increased to 809. The figures for 1922 showed a decided shift away from the old type of schedule in favor of what is rapidly becoming known as a *unit schedule*. At least 75 cities had either revised their old schedules or adopted unit schedules for the first time. The unit schedule bases its increases to teachers upon (a) academic and professional training, (b) length of service, and (c) merit.

Should increases be based in part upon length of service?—Yes, all schedules have recognized this principle. Furthermore, increases should continue as long as additional service renders one increasingly efficient. No one knows just how long this is. It is generally agreed, however, that service beyond a certain number of years does not continue to improve the quality of teaching, at least at the same rate. There is no proof of this statement one way or the other. The writer believes in continuing increases at a decreasing rate even up to old age, assuming, of course, continued efficiency.

What recognition should be given to service outside of the system?—While practice tends to be provincial in this matter, there is a growing tendency to consider service in one school system the equal of service in another system. This seems sound and equitable. What difference does it make where the experience was secured? Even country-school experience, especially if intelligently supervised and directed, may be approved. If it is good for a teacher to change towns she should be able to do so without penalty, unless she is inefficient or incompetent.

What recognition should be given to service outside of teaching?—None, unless that outside service is directly related to the teaching concerned. This may be true in the case of commercial, agricultural, and industrial subjects. Even then it may be doubted if one year of practice is the equivalent of one year of teaching. Possibly a good rule would be to require, as in the laboratory part of the study of a science, three years of practical experience for one year of teaching. Hardly any of the schedules have as yet recognized experience outside of teaching. However, modifications of this nature will occur in the future. The following has been proposed in one city to care for this problem: (a) Credit will be allowed for experience outside of the teaching profession, which is approved by the educational committee upon the recommendations of the head of the department, the principal, and the superintendent of schools. (b) In order to receive credit, such practical experience must be closely related to the subject or subjects subsequently taught and, in the judgment of the majority concerned, be of direct value in equipping the person for such teaching. (c) Each year of such approved practical experience, up to and including ten years of such experience, shall count not to exceed the equivalent of one-half year of teaching experience.

Should increases be based upon promotional examinations and certificates?—No, unless these are equated in terms of academic and professional courses in recognized institutions for teaching training. Why should each school system attempt to set up a separate accrediting agency when higher educational institutions

have been doing this for years? Much less attention should be paid to examinations and certificates, and much more attention should be given to regular courses in established institutions.

Should increases be based upon academic and professional training?—Yes, such training costs money and represents an investment of capital that should produce returns. The amount of training that is recognized in the schedule should be unlimited and the kind should not be too narrowly restricted. Anything that improves one's personal and professional culture and skill in a substantial degree should be recognized.

Should merit be recognized?—Yes, wherever and whenever it is possible to measure merit accurately. Is it now possible to measure the merit of a teacher? This opens up a whole new field of research, especially if we attempt to measure merit objectively. Courtis says, "The best proof of the pudding is in the eating, and the best proof of ability in a teacher is the effect produced by her teaching." We are all perfectly certain that we can tell a good teacher when we see one and yet we all make mistakes in our estimates. The teacher's scholastic record and her promotional record, together with that vague thing known as personality, are the essential factors used in evaluating her. Only after a series of long continued tests would it be possible to measure the effects produced by her teaching. Teachers now feel that they are too frequently misjudged by their principals, supervisors, and superintendents. This leads them to distrust all present methods of measuring merit.

Should increases be granted for travel?—As far as the writer is aware, no university or teacher-training institution gives any credit for travel unless such travel is highly organized, supervised, and directed by regular professors. A school system should not, as a rule, attempt to set up a separate accrediting agency but should use the accrediting agencies already in existence. Undoubtedly, travel when properly directed, is highly educational; but so too is general reading, concert attendance, etc. Sometimes travel rejuvenates a tired teacher more than academic or professional study. However, teachers as a group are still

woefully lacking in the personal and professional culture so much needed for the highest teaching service.

Should the distinction between sexes be recognized?—The theoretical answer would be "no;" but, practically, it is not yet possible in most places to wipe out the distinctions in salaries that exist between men and women. Such distinctions have not been wiped out in other trades, and it is difficult for education to go ahead of other trades, thus setting up a basis that does not operate elsewhere. The movement for "equal-pay-for-equal-work" is as old as society itself. In teaching, it first attracted national attention in 1906 when it was made the chief issue of educational legislation in New York City. Finally, in 1912, the principle of "paying the position" regardless of sex was adopted. Similar campaigns were successful in San Francisco and Los Angeles; and in 1916 the legislature of Oregon adopted the principle as a state measure. The phrase "equal-pay-for-equal-service" is misleading. What is really meant is equal pay for equal merit, equal length of service, equal training, and equal work. Where it is impossible to wipe out the distinction, it may be possible to justify the difference in terms of necessity, supply and demand, or the difference in hours of work, or duties undertaken.

The Salary Schedule, Rockford, Illinois

The schedule adopted at Rockford, Illinois, January, 1923, under the writer's direction, has certain features never before fully recognized. The distinction between grade, junior, high-school and senior high-school teachers is entirely ignored. Teachers are paid according to length of service and amount of training. No limit is placed on the amount of service recognized, and no attempt is made to measure merit on account of the diversity of ideas as to how it should be measured. This is a limitation of the schedule that can easily be amended. A suggested amendment would read like this: All *A* teachers are to be advanced one-fourth or one-third faster than this schedule; all *B* teachers are to advance in accordance with this schedule;

TRAINING AND EXPERIENCE SCHEDULE, ROCKFORD, ILLINOIS

Experience	Years Training												
	2	2½	2⅔	3	3⅓	3⅔	4	4½	4⅔	5	5½	5⅔	6
0	1000	1050	1100	1150	1200	1250	1300	1350	1400	1450	1500	1550	1600
1	1050	1100	1150	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650
2	1100	1150	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700
3	1150	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700	1750
4	1200	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700	1750	1800
5	1250	1300	1350	1400	1450	1500	1550	1600	1650	1700	1750	1800	1850
6	1300	1350	1400	1450	1500	1550	1600	1650	1700	1750	1800	1850	1900
7	1350	1400	1450	1500	1550	1600	1650	1700	1750	1800	1850	1900	1950
8	1400	1450	1500	1550	1600	1650	1700	1750	1800	1850	1900	1950	2000
9	1450	1500	1550	1600	1650	1700	1750	1800	1850	1900	1950	2000	2050
10	1500	1550	1600	1650	1700	1750	1800	1850	1900	1950	2000	2050	2100
11	1520	1570	1620	1670	1720	1770	1820	1870	1920	1970	2020	2070	2120
12	1540	1590	1640	1690	1740	1790	1840	1890	1940	1990	2040	2090	2140
13	1560	1610	1660	1710	1760	1810	1860	1910	1960	2010	2060	2110	2160
14	1580	1630	1680	1730	1780	1830	1880	1930	1980	2030	2080	2130	2180
15	1600	1650	1700	1750	1800	1850	1900	1950	2000	2050	2100	2150	2200
16	1610	1660	1710	1760	1810	1860	1910	1960	2010	2060	2110	2160	2210
17	1620	1670	1720	1770	1820	1870	1920	1970	2020	2070	2120	2170	2220
18	1630	1680	1730	1780	1830	1880	1930	1980	2030	2080	2130	2180	2230
19	1640	1690	1740	1790	1840	1890	1940	1990	2040	2090	2140	2190	2240
20	1650	1700	1750	1800	1850	1900	1950	2000	2050	2100	2150	2200	2250
21–25	1650	1700	1750	1800	1850	1900	1950	2000	2050	2100	2150	2200	2250
26–30	1700	1750	1800	1850	1900	1950	2000	2050	2100	2150	2200	2250	2300
31–35	1750	1800	1850	1900	1950	2000	2050	2100	2150	2200	2250	2300	2350
36–40	1800	1850	1900	1950	2000	2050	2100	2150	2200	2250	2300	2350	2400
41–45	1850	1900	1950	2000	2050	2100	2150	2200	2250	2300	2350	2400	2450
46–50	1900	1950	2000	2050	2100	2150	2200	2250	2300	2350	2400	2450	2500

($50 increase for each additional five-year period.)

all *C* teachers are to stand still until they get a rating of *B* or better; and all *D* teachers are to be dropped or put on probation.

Comments on the Rockford Schedule.—A minimum of $1,000 will be paid to a teacher without experience whose training is equivalent to normal-school graduation. For each additional year of experience, $50 will be added, up to and including ten years. For the eleventh to the fifteenth years inclusive, an annual increase of $20 per year will be given. From the sixteenth to the twentieth years inclusive, an increase of $10 per year will be given. Beyond the twentieth year limit an increase of $50 for each five-year period will be allowed.

For each one-third year of training (ten semester hours or equivalent) beyond normal-school graduation or equivalent, $50 will be added to the minimum salary, up to and including a total of four years of such training. Credit for training is cumulative and each half-year salaries will be adjusted accordingly.

Teachers whose present salaries are below this schedule will be given increases not exceeding $100 per year until they reach the schedule. However, teachers of twenty to thirty years' experience will be raised to their schedule in two equal annual raises, and teachers of thirty or more years' experience will be raised to their schedule immediately. All such teachers shall participate in further advances in accordance with the schedule. Only ten years of experience outside of Rockford, however, shall be credited toward salary increases on account of experience. Teachers who in 1922-23 were paid more than such a schedule would permit were not to have their salaries reduced during that year.

Conclusion

A word should be said in conclusion. The writer believes that more attention should be paid to the principles upon which salary schedules are built. The amounts are important; but unless sound principles are recognized and applied, the schedule will fail or at least prove unsatisfactory. This paper is the result of the writer's experience and he will welcome the criticisms and suggestions of others. Only thus can we reach the right solution to this most baffling problem.

THE USE OF THE CORRELATION GRAPH WITH HALF-SIGMA CLASS INTERVALS

GERTRUDE RAND
Bryn Mawr College

In an article entitled "Combining the Results of Several Tests: A Study in Statistical Methods" Woodworth[1] recommends that for comparing or combining the records of different individuals in the same test or of the same individual in different tests, all test records be reduced to fractions or multiples of some measure of variability of the distributions in question. These "reduced measures" thus become pure numbers which are comparable in magnitude provided the distributions under comparison are of the "normal" type or if all belong to any one type. Woodworth states (p. 104) that "although the labor involved" in obtaining these reduced measures is considerable, "once this labor is performed the operation of computing coefficients of correlation is considerably simplified." If, for example, the standard deviation (σ) is taken as the unit of deviation and the various test records are reduced to positive or negative fractions or multiples of σ (that is, $\frac{d}{\sigma}$ where d is the deviation from the mean), then the Pearson product-moment formula for the computation of the coefficient of correlation,

$$r = \frac{\Sigma\, xy}{N\, \sigma_1\, \sigma_2},$$

becomes simply the average product of the reduced measures for the two arrays under consideration. In other words,

$$r = \frac{\Sigma\, ab}{N}$$

in which a is the reduced measure for an individual in the x-series; b, the reduced measure for the individual in the y-series; and N, the number of cases.

[1] Woodworth, R. S. *Psychological Review*. 19:97-123. March, 1912.

I doubt, however, whether it is generally realized how much the labor of computing r is reduced if Woodworth's method is used in conjunction with a correlation graph or double entry table the class ranges of which are always chosen to cover $0.5\,\sigma$ of the total distribution; and if it be assumed, as is commonly done,[2] that all the values within any class-interval are distributed uniformly throughout the interval and may without serious inaccuracy be represented by the value of the mid-point. When this type of double entry table is used, the individual measures are entered into the table directly in their reduced form and the arithmetic of the computation of r becomes almost entirely the arithmetic of unit and half-unit values ranging from 0.5 to 3 instead of the original raw scores which usually range to 200, often to 400 or 500. As one of the additional advantages of the plan proposed, it may be mentioned that the worker is forced to include the correlation graph as a necessary adjunct to the determination of r, and thus to demonstrate whether or not linearity of relationship is present in the data correlated—a procedure too frequently neglected.

The steps in this simplified procedure are as follows. For illustration reference may be had to Figures 1 and 2.

(1) The mean and standard deviation are computed for both the x- and y-series to be correlated. Any of the approved short methods which involve the use of class-intervals may be used in making these computations. Since, however, the class-intervals to be used in the correlation graph are in terms of half-*sigma* values and cannot therefore be determined in advance of the computation of the means and the standard deviations, I have found it a simpler procedure to compute the means by the long

---

[2] On the use of class-intervals and the assumptions underlying the representation of all values falling within the interval by the mid-point, see discussions in:

Rugg, H. O. *Statistical Methods Applied to Education.* New York, Houghton Mifflin Company, 1917. p. 83 ff.

Thorndike, E. L. *An Introduction to the Theory of Mental and Social Measurements.* 2d Ed. New York, Teachers College, Columbia University, 1919. pp. 42-60.

Yule, G. U. *An Introduction to the Theory of Statistics.* London, C. Griffin and Company, 1911. p. 79 ff.

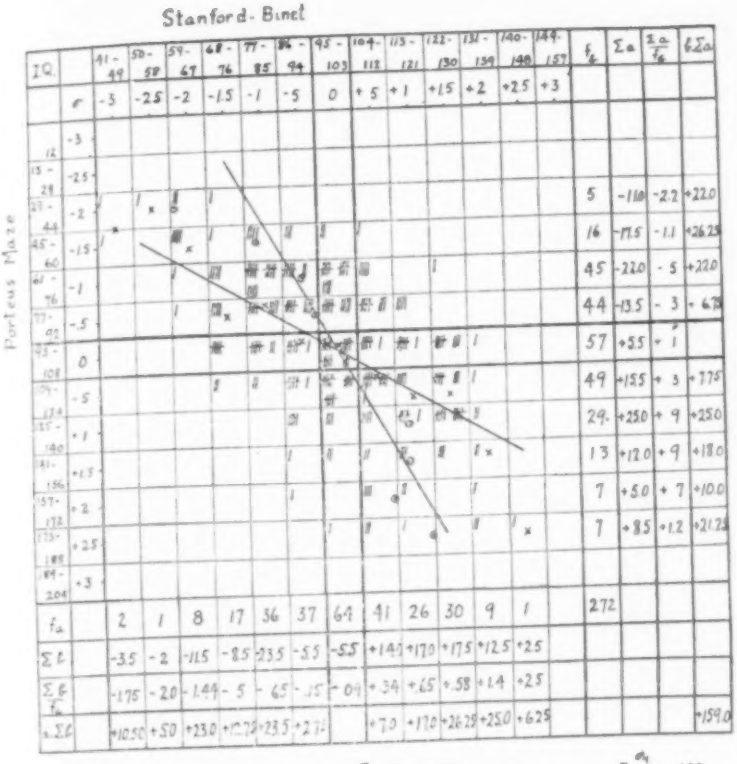

FIGURE 1. CORRELATION BETWEEN INTELLIGENCE QUOTIENTS AS DETERMINED BY THE STANFORD-BINET TEST AND INTELLIGENCE QUOTIENTS AS OBTAINED BY THE PORTEUS MAZE TEST

method using an adding machine, and the standard deviations by Ruml's formula,[3]

$$\sigma = \sqrt{\frac{\Sigma\,(\text{score}^2)}{N} - \text{mean}^2},$$

using the adding machine and a table of squares. Inasmuch as

[3] Ruml, Beardsley. "On the Computation of the Standard Deviation." *Psychological Bulletin.* 13:444-46. November, 1916.

the individual deviations from the mean do not enter into the later computations of r, no advantage is gained from including them in the process of determining σ.

(2) The second step is the laying out of a double entry web such that the central class-interval for each coordinate includes those values of the variable which range from $+0.25\,\sigma$ to $-0.25\,\sigma$; and the class-intervals on either side, the values which range successively from $\pm 0.25\,\sigma$ to $\pm 0.75\,\sigma$; $\pm 0.75\,\sigma$ to $\pm 1.25\,\sigma$; $\pm 1.25\,\sigma$ to $\pm 1.75\,\sigma$; $\pm 1.75\,\sigma$ to $\pm 2.25\,\sigma$; $\pm 2.25\,\sigma$ to $\pm 2.75\,\sigma$; $\pm 2.75\,\sigma$ to $\pm 3.25\,\sigma$. The mid-values of the class-intervals are then, in terms of σ, 0, .5, 1, 1.5, 2, 2.5, and 3. For convenience in tabulating the individual test records, the score values limiting each half-*sigma* interval are listed, in Figures 1 and 2, in the row above the σ designations for the x-array; and in the column to the left, for the y-array. For convenience in computation, the σ designations are made for the mid-values of the intervals only.

(3) The individual pairs of measures are then entered into the appropriate squares of the table, and from this point on they are treated as reduced measures, not score values. The assumption is made, as stated above, that the reduced measures are concentrated around the mid-value of the class-interval and can be represented by this value for computation purposes without serious inaccuracy.[4] Each entry may now be read directly from the table either as the $\frac{x}{\sigma_1}$ of Pearson's formula and the a of Woodworth's; or as the $\frac{y}{\sigma_2}$ of Pearson's and the b of Woodworth's.

(4) Columns are arranged at the right of the diagram for the following entries for each row of measures: (1) for the frequency of occurrence of the measures in each row (f_b), an

[4] I have computed r by both the long method and the method described in this paper for a great many cases and have not as yet found a difference in result by the two methods greater than $\pm$ 0.015. Theoretically r as determined above should be slightly too high. That is, the theoretical error which enters into all computations by the method of grouping will in this case affect only the numerator of the fraction. In the usual short method of computing r the error occurs in both numerator and denominator. Whether it can be considered that this operates to cancel the error is open to question.

inspection of which will show the form of the distribution of the y-variates; (2) for Σa; (3) for $\frac{\Sigma a}{f_b}$, or the mean a for each row; and (4) for $b.\Sigma a$. Since all of the a's occurring in one row are associated with the same b, the product ab for all values in a row can be made in one operation.

(5) Similarly, rows are arranged at the bottom of the diagram to give (1) the frequency of occurrence of the measures in each column (f_a), thus showing the form of distribution of the x-variates; (2) Σb; (3) $\frac{\Sigma b}{f_a}$, or the mean b for each column; and (4) $a.\Sigma b$.

(6) The mean values of a for the rows are entered into the correlation graph in their appropriate places where they are designated by small circles; also the mean values of b for the columns, designated by small crosses. Inspectional regression lines are then provisionally drawn through the means of the a's and the means of the b's to establish the presence of linearity of relationship before the last step in the computation of r is taken.

(7) The entries in the column headed $b.\Sigma a$ are summed and divided by the total number of cases N. The result gives r. As a check on the accuracy of the work the entries in the row headed $a.\Sigma b$ may also be summed and divided by N.

(8) To complete the comparison of the two traits, the regression coefficients are determined; also the most probable values of y for any two values of x and the most probable values of x for any two values of y. These will furnish sufficient data for the exact location of the regression lines.

The time saved in this procedure as compared with that of Woodworth lies (a) in obtaining the individual reduced measures simply by entering them into their appropriate class-interval which is designated by its score value and by either its *sigma*-range or its *sigma*-mid-point; and (b) in treating the reduced measures for computation purposes in groups which are represented by the mid-value of the class interval. The method is fully as quick as if not quicker than either the short method de-

scribed by Yule[5] and by Rugg[6] or that of Otis;[7] and is, I believe, simpler in conception and procedure than either of these. It is to be preferred to them, however, mainly because of the many advantages that are to be derived from having the class-intervals for the two arrays in comparable units. That is, by this method all of the advantages of the reduced measure as a unit are retained and in addition the procedure used in computing r is greatly simplified and shortened.

Two examples of data plotted on the correlation graph described above and of r computed by this procedure are given in Figures 1 and 2. Figure 1 consists of test records of all the first-grade children (*272* cases) of a suburban public school district near Philadelphia. The district comprises six schools, two of which give approximately a "normal" distribution of test records; two a distribution skewed to the upper end of the curve; and the remaining two, a distribution skewed to the lower end. The tests correlated are the Stanford-Binet individual test (x-series), the records for which were carefully taken with wide-range testing, and the Porteus Maze individual test (y-series). The test records are expressed in the quotient form in both cases. The mean value of the quotient for the Stanford-Binet test is 99, and the value of σ is 18;[8] for the Porteus Maze test the mean value is 101 and the σ is 32. In the column headed f_b is given the frequency for each half-*sigma* range for the Porteus Maze test, and in the row headed f_a, the frequency for the Stanford-Binet. Any row of the table shows the scatter of quotients of the Stanford-Binet for any half-*sigma* range of the Porteus quotients; similarly any column shows the scatter of quotients of the Porteus Maze for any half-*sigma* range of the Stanford-Binet quotients. The coefficient of correlation is

[5] Yule, G. U. *Op. cit.*, pp. 181-88.

[6] Rugg, H. O. *Op. cit.*, pp. 260-70.

[7] Otis, A. S. "The Otis Correlation Chart." *Journal of Educational Research*, 8:440-48, December, 1923.

[8] The σ for the entire grade is, of course, considerably larger than that given by any one age group within the grade. For the first-grade children, whose chronological age ranged from 6.5 to 7.5 (103 cases), the Stanford-Binet mean quotient is 101.4 and the σ, 13.5; for the Porteus Mazes, the mean quotient is 102.4, and the σ, 30.

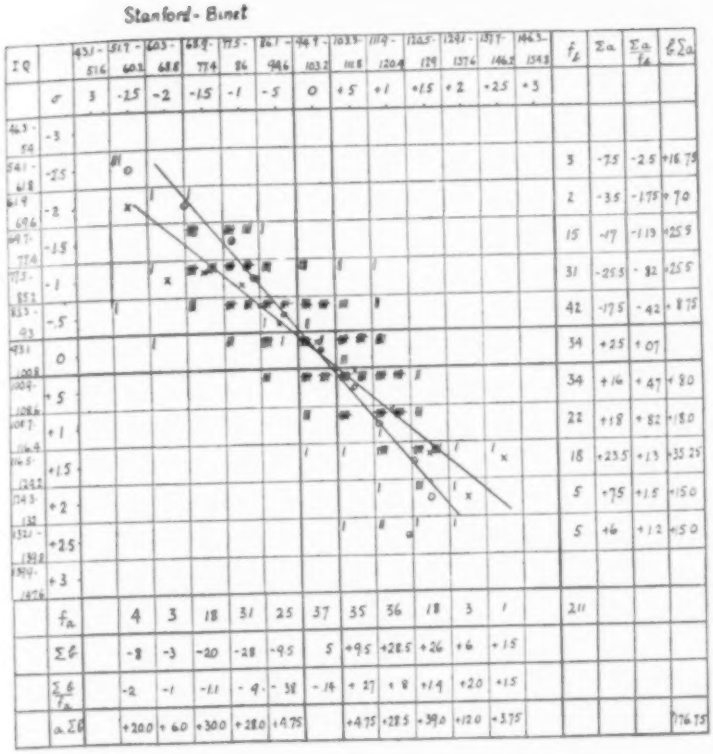

FIGURE 2. CORRELATION BETWEEN INTELLIGENCE QUOTIENTS AS DETERMINED BY THE STANFORD-BINET TEST AND INTELLIGENCE QUOTIENTS AS DETERMINED BY THE DEARBORN GROUP INTELLIGENCE TEST

$+0.58 \pm 0.027$. The regression coefficient of y on x is 1.03; of x on y, 0.33.

The second example, Figure 2, consists of test records of first-grade children (211 cases) of the same school district taken during another school year. One of the two schools whose distribution of Stanford-Binet intelligence quotients was skewed to the upper values and one of the two schools where this dis-

tribution was skewed to the lower values were omitted from this comparison. The tests correlated are the Stanford-Binet individual test (x-array) and the Dearborn Group Intelligence Test Series I (y-array). The group test was given during the latter part of the first semester of the school year. The records are expressed as intelligence quotients for both tests. The mean value of the quotient for the Stanford-Binet test is 99 and the σ is 17.2; for the Dearborn test the mean value of the quotient is 97 and the σ is 15.5. The coefficient of correlation is $+0.84 \pm 0.014$. The regression coefficient of y on x is 0.754; of x on y is 0.93.[9]

As illustrative of the many uses which may be made of this type of correlation graph in the evaluation and interpretation of test data, the following may be mentioned:

(1) The score-values used within the class-intervals for the two arrays are directly comparable, assuming that the distributions approach the "normal" or a similar type. That is, the σ of each interval is a pure number and the individual cases located in each square of the graph can be directly compared in terms of their units of deviation from the means of the arrays. The advantage of this when means and σ's differ in kind and magnitude is obvious. In the two examples given above, the first (intelligence quotients, Stanford-Binet, correlated with intelligence

[9] Fourteen records could not be included in the above correlation because the Dearborn I.Q. could not be obtained. In these cases the scores obtained in this test fell below 7, the standard for age 5 which is the lowest value given by Dearborn in his table of standards. Only four of these cases had a mental age in the Stanford-Binet test above 5. In order to give all the comparative data obtained for the two tests, the chronological age, mental age, and I.Q. (Stanford-Binet) and the score in points (Dearborn) for the fourteen cases are appended.

Number of Case	Chronological Age	Stanford-Binet		Dearborn Score
		M. A.	I. Q.	
1	6.17	5.42	88	6
2	6.17	5.00	81	5
3	6.83	4.92	72	5
4	6.75	4.58	68	5
5	11.58	4.92	42	4
6	6.17	4.67	76	4
7	7.58	5.33	70	4
8	6.25	4.67	75	3
9	7.50	4.33	58	2
10	6.50	5.33	82	2
11	6.17	3.83	62	1
12	6.42	3.00	47	1
13	6.50	5.42	83	1
14	6.83	4.50	66	0

quotients, Porteus Maze) was chosen because of the dissimilarity of the σ's for the two tests; and the second (intelligence quotients, Stanford-Binet, correlated with intelligence quotients, Dearborn Group Intelligence Test), because of the approximate similarity of the two σ's. For all three tests the records are expressed in terms of the intelligence quotients and in all three the value of the mean quotient lies close to 100. Nevertheless the intelligence quotients above and below 100 have a slightly dissimilar quantitative significance in the Stanford-Binet and the Dearborn tests, and a very dissimilar quantitative significance for the Porteus Maze test, as is seen by comparing the score values limiting the same *sigma*-range for each test. This demonstration that the intelligence quotient as a general concept has no quantitative significance from test to test except at the mean is too obvious to require further emphasis.

(2) The individual cases can be directly compared also in terms of their units of displacement from the regression lines. This comparison shows at a glance which are the cases that deviate widely from the general trend of x-y relationship and should therefore be given special study as to the cause of the deviation or perhaps individual treatment and recommendation. For example, in the correlation graph given in Figure 1 of the Stanford-Binet and Porteus Maze quotients, cases far removed from the regression lines locate the children of wide divergence in intelligence as rated by the Stanford-Binet test and in intelligence combined with the character traits stressed in the Porteus ratings. In other words, the children obtained ratings in these character traits which are higher or lower by known amounts of the *sigma* unit than the average rating of children of the same Stanford-Binet standing. Often, too, this type of case study will suggest causes of the lack of agreement between the results of the tests under comparison. For example, it has frequently been my experience to find that when the results of a non-verbal group test, such as the Pressey Primer or the Dearborn Series I, are correlated with those of the Stanford-Binet verbal individual test for young children, the cases which deviate widely from the

regression line toward lower unit ratings in the Stanford-Binet test are children of foreign parentage who through lack of fluency in English are at a disadvantage in this type of test; and on the other hand the cases which deviate from the regression line toward lower unit ratings in the group test are children who were characterized by their teachers or were observed and noted during the administration of the individual test to be shy or slow or nervous, and therefore at a disadvantage in a group test. Comparisons of this type would not have been possible had the class-intervals been equal ranges of test score as they are in the usual correlation graph. They could have been made in such cases only if the means and the σ's of the two arrays had been approximately equal in value, which rarely happens.

(3) When the correlation graph is laid off in half-*sigma* intervals, the row and the column of frequencies occurring within each class-interval furnish the data directly for histogram plots of half-*sigma* intervals. A comparison can then be readily made between the percent of the total frequency falling within each class-interval with the percent for theoretically "normal" or other known type of distribution.

And (4) since the σ for each distribution is represented by an equal distance along both the x and the y axes, the historically interesting graphic method of Galton for determining r can be applied as a further check on the accuracy of the computed r. For this method, a horizontal line from any point A on the vertical median line of the graph is drawn so that it cuts the line of perfect correlation (the 45° axis of the graph) at B and the line of the means of the columns at C. Then the ratio AB : AC is equal to r.

THE INFLUENCE OF LATIN ON THE SPELLING OF ENGLISH WORDS

WARREN W. COXE
Bureau of Educational Measurements,
State Department of Education
Albany

The present study is part of a larger investigation conducted by the American Classical League. This larger investigation aims to determine what should be the objectives in the teaching of Latin in the secondary schools, and what methods and materials are most effective in attaining these objectives.

The phase of the whole investigation herein reported is the influence which Latin has and can have upon the spelling of English words. The study falls naturally into two divisions: (1) the extent to which Latin, as now taught, is improving the spelling of English words and (2) the best methods and material which can be used to produce a maximum improvement.

Accordingly the investigation follows two lines: a general survey, in which the relative gains of Latin and of non-Latin pupils in spelling were measured; and a controlled experiment, in which the conditions of transfer were studied more in detail.

THE GENERAL SURVEY

In order to determine to what extent Latin, as now taught, influences the spelling of English words, 58 schools located in different parts of the United States were asked to cooperate in a general survey. The Buckingham-Coxe Spelling Scale, which was prepared especially for this study, was given in these schools to pupils in the English classes of the year in which Latin was begun. Thus, pupils who were beginning Latin, as well as those who were not taking Latin were tested. A second form of the test was given at the middle of the year to measure the gain made during the first semester, and a third form was given at the end of the year to measure gains made during the year. The Buckingham-Coxe Spelling Scale consists of fifty words, twenty-five of which were derived from non-Latin sources and twenty-

five from Latin words studied at some time during the high-school course. Of the latter twenty-five, ten were derived from Latin words studied in the first year. The spelling difficulties of the words in this scale were known and all three forms were made of equal difficulty.

After the papers from the 58 schools had been corrected, they were so tabulated as to show the gains made during the semester and during the year. In order to compare the spelling gains made by the Latin pupils with the gains made by the non-Latin pupils, it was necessary to equate the two groups on the basis of initial spelling ability. When this was done, the Latin pupils showed an average gain for the year of 3.6 words; the non-Latin pupils, a gain of 2.6 words. There is thus an advantage for the Latin pupils of one word. These statements refer to the twenty-five words derived from Latin studied somewhere in the high-school course. A comparison of the gains made on Latin words for the semester and for the year shows that most of the gain was made during the first semester. On words of non-Latin origin, the Latin pupils gained 0.2 of a word, while the non-Latin pupils gained 0.1 of a word.

The conclusion, therefore, is warranted that in general, Latin as now taught does improve the spelling of English words of Latin derivation, but does not assist in the spelling of words of non-Latin origin. Nearly all of this improvement takes place in the first semester.

The Controlled Experiment

Although the results of the general survey show that Latin does assist in the spelling of English words, they do not show to what extent the gains depended upon the method. Accordingly a controlled experiment was carried out to determine whether the amount of gains could be increased by conscious efforts to relate Latin to English spelling, and, if so, what methods produced the greatest gains.

Several groups of classes were organized, each following a slightly different procedure. The groups were as follows:

Latin Control Group (Group I).—This group was composed of pupils beginning Latin. They followed the methods usually prescribed by the school, except that the teachers were cautioned not to relate Latin to the spelling of English words.

Latin Experimental Alpha Group (Group II).—The pupils of this group were of the same grade as those in the preceding group. This group followed a definite teaching method and used definite practice material. The method consisted in pointing out the similarity between the Latin word and its derivative. The teachers were cautioned not to develop any rules or principles, which would show the relation between the Latin word and its derivatives. The material used consisted of 153 English words, all of which were derived from Latin found in the pupil's textbook.

Latin Experimental Beta Group (Group III).—This group used definite methods and material. The material was the same as that of the preceding group, but the methods involved the development of principles. The list of principles follows:

1. Original double consonants are regularly preserved in derivatives (except at the ends of compounds). *Terra* has two *r*'s. Therefore the derivative terrestrial has two *r*'s.
2. The "obscure" vowel follows the original Latin. *Tempore* is spelled with an *o*. Therefore temporal is spelled with an *o*. (Note to Rule 2: When the conjugations have been developed, the relation of the obscure vowel to the conjugation can be pointed out, e.g., in portable the obscure vowel *a* is the stem vowel of the first conjugation.)
3. Many consonants and combinations of consonants whose pronunciation has changed are preserved in English. *Discipulus* is spelled with "sc." Therefore disciple preserves "sc."
4. When a prefix ending in a consonant ("ad," "con," "in," "ex," "ob," "dis," "sub") is prefixed to a word beginning with a consonant, the first consonant is assimilated, if possible, to the second, and double consonants are produced in the derivative. Affiliate is derived from the prefix *ad* and *filius*. Therefore affiliate has two *f*'s.
5. Initial *s* after "ex" is lost. *Exspecto* has an *s* after "ex." In the derivative expect the *s* is lost.
6. Forms in "ant," "ance" and "ent," "ence" are present active participles and follow the original Latin. Words of the first conjugation end in "ant," all others in "ent"; "io" verbs end in "ient." *Porto* belongs to the first conjugation. Therefore important is spelled with "ant." Exception: Some words which should end in "ent" become nouns and are spelled "ant." When the original part of speech, i.e., adjective, is retained, the original spelling is also retained, e.g., confidant (n), confident (adj.).

7. The variation between "sion" and "tion" is determined by the fourth principal part of the Latin verb from which the word is derived. The last principal part of *peto* is *petitus*. Therefore petition is spelled "tion."

English (non-Latin) Control Group (Group IV).—These classes did not take Latin and no special effort was made to teach spelling other than that required by the school in which the class was taught.

English (non-Latin) Experimental Group (Group V).—These classes used the same material as that prescribed for the two Latin experimental groups. The method which was employed did not presuppose any Latin, but followed the best current practice in the teaching of spelling.

All five groups were given the Buckingham-Coxe Spelling Scale at the beginning of the experiment in November, 1922, again in February, and in May, 1923. In addition, a supplementary test was used consisting of 100 words derived from Latin words usually studied in the first year. This test was known as the Ungraded Spelling Test and was given at the beginning of the experiment and repeated at the end. The difficulties of the words in the Ungraded Spelling Test were not known, but obviously easy words were not included.[1]

The gains made by Group I, composed of Latin classes in which no effort was made to relate Latin to English spelling were compared with those made by Group IV, composed of non-Latin English classes, in which no special method in the teaching of spelling was used. The pupils of these two groups were paired on the basis of initial spelling ability. On words derived from Latin words studied somewhere in the high school, the seventh-grade pupils gained 6.3 words during the year, while the English pupils gained 5.0 words. In the ninth grade, similar figures are 3.0 and 2.7. Thus it appears that Latin, even though not related by the teacher to the spelling of English words, did help to improve their spelling.

Ten of the words of our spelling scale were derived from Latin words studied in the first year. The superior gain of Latin

[1] For a more complete description of the method of the controlled experiment, see Coxe, W. W. "A Controlled Experiment to Determine the Extent to Which Latin Tests Can Function in the Spelling of English Words." *Journal of Educational Research.* 7:244-47, March, 1923.

pupils on these words is greater than on the Latin words previously mentioned.

On non-Latin words, the results are reversed. The Latin pupils tend to lose, while the non-Latin pupils make slight gains. This indicates that Latin interferes slightly with the spelling of non-Latin words. It may be that the pupil who studies Latin sets up a sort of scheme for spelling which he tries to apply to all words. This leads to incorrect spellings of words which do not follow the Latin scheme.

It is also interesting to know in what part of the year the gains were made. In the seventh grade, the superiority of the Latin pupils tended to increase constantly throughout the year. In the ninth grade, considerable gain was made in the first semester, but almost none in the second. We have reports from several teachers to the effect that ninth-grade Latin was congested during the latter part of the year, due to the need of preparing for *Cæsar*. The seventh-grade classes did not feel this strain because this grade attempted to cover only the work which is ordinarily covered in the first semester of the ninth grade.

Even though measurable gains resulted when no special effort was put forth to attain them, we may well ask how much additional gain can be obtained when special methods are used. Two special methods were used which have been described under the procedures of Groups II and III. The classes of Group II merely noted the similarity in spelling between Latin and English words. The classes of Group III developed definite principles of similarity. We compared the gains made by these two methods with the gains made by Latin pupils who had used no special method (Group I). On words derived from Latin words studied during the high-school course, the pupils of Group II showed a little superiority over the pupils of the control group (Group I). The pupils of Group III had a distinct superiority. The pupils of the control group gained 6.7 words while the pupils of Group III gained 9.1 words. Almost the same results are found when we compare these groups with respect to words derived from Latin studied in the first year. Twenty-five of the

words which were in the practice material covered by Groups II and III were also in the Ungraded Spelling Test. On these words, Groups II and III showed a greater superiority over the control group than on Latin words in general. It is particularly interesting to note that the method used in Group II tends to interfere more with the spelling of words of non-Latin origin than the method used in Group III. While this fact is negative in character, it indicates that care must be taken in teaching Latin so that it will not cause a loss of any kind. The use of principles seems to guarantee that the ability to spell words of non-Latin origin will not be impaired, and also that the ability to spell words of Latin origin will be increased.

The particularly valuable result in the above comparison is that special methods can produce a greatly increased ability to spell words of Latin origin.

If we compare these special methods directly as to their relative efficiency, the method used in Group III is apparently more effective than that used in Group II. On words derived from Latin words studied during the high-school course the pupils of Group II gain 8.4 words during the year as compared with 10.2 words, gained by the pupils of Group III. On words of non-Latin origin the pupils of Group III lose less than the pupils of Group II. Evidently as a means for producing the greatest possible gains on words of Latin origin and the least possible interference with words of non-Latin origin, the method used in Group III is superior.

An interesting question arises concerning the influence of the methods used in Groups II and III on individual words. It seemed possible that the development of principles might correct certain spelling errors and leave others uncorrected. It might be that only the errors covered by principles would be corrected. Although further study of this phase of the matter is being made, the analysis so far shows that errors not covered by the rules were corrected. The training gained through the use of principles seemed to lead to a critical examination of the whole word

rather than to any part. The method used by Group II did not do this so effectively.

Some difficulty was experienced by the teachers of the seventh grade in teaching the principles involved in the method of Group III. The language of the rules seemed to be difficult for the pupils and some of the rules could not be taught satisfactorily until certain phases of the grammar had been covered. In spite of this difficulty such method produced as good results in the seventh grade as it did in the ninth grade. In both grades it was distinctly superior to the method which did not use rules.

The spelling gains made by the two experimental Latin groups were compared with those made by an experimental non-Latin English group which was given the practice material, but used a method of teaching spelling which involved the best experimental results available. The results show that the experimental English classes make spelling gains superior to those made by the Latin classes which merely point out the relationship of the Latin to English spelling. On the other hand, their results are somewhat inferior to those of the Latin classes in which principles were developed. These results apply to Latin words in general. On words derived from Latin studied in the first year, both the Latin experimental groups show superior gains. On words which were included in the practice material, the non-Latin English experimental group is very distinctly inferior in its gains to the two Latin experimental groups. On non-Latin words, all the Latin groups, including both experimental groups, are inferior in gains to the English experimental group.

Another kind of comparison is possible. We may consider that we have three types of words of Latin origin: (1) words derived from Latin words studied somewhere in the high school, (2) words derived from Latin words studied in the first year, and (3) words which were included in the practice material. Both types of experimental Latin classes show an increasing amount of gain as we go from words found in high-school Latin to words of first-year Latin origin and thence to words included in the practice material. These methods are most effective in

aiding English spelling when the words included in the test are also the words which were used in the practice material.

Summary

It is desirable that these results be related in such a way that we may get a bird's-eye-view of the whole situation. This has been done in Figures 1, 2, 3, and 4. Our interest is not primarily in the amount of gain made by each of our groups, but rather in the amount by which these gains exceeded those customarily secured. We have assumed that the gains made by the English (non-Latin) group (Group IV) represent the customary situation, and we have calculated the amount by which the gains made by Groups I, II, III, and V exceed the gain made by Group IV.

Figure 1 shows the influence of the different methods of instruction upon words derived from Latin studied during the high-school course. The special method used in the non-Latin classes (Group V) produces a greater advantage in amount of gain than the Latin instruction which did not seek to improve spelling (Group I) and also greater than the Latin instruction which related Latin to English spelling but did not develop rules (Group II). The group of Latin classes which developed rules (Group III) shows an advantage over all other methods.

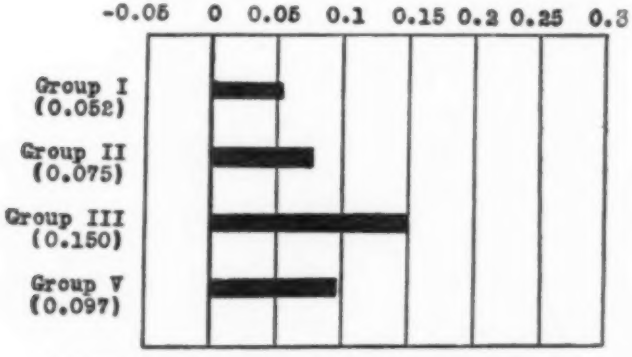

FIGURE 1. LATIN WORDS. SUPERIORITY PER PUPIL PER WORD OF EACH GROUP OVER GROUP IV (ENGLISH CONTROL)—GRADE VII

The relative influence of the various types of instruction on words derived from Latin which has been studied is shown in Figure 2. The three groups of Latin classes hold the same relative position they did in the preceding figure. The bars are longer, however, indicating that these groups did better on words of this type than on the words represented in Figure 1. The non-Latin group which used a special method, shows the smallest excess of gain.

All three Latin groups show in Figure 3 their maximum improvement over Group IV on the words included in the practice

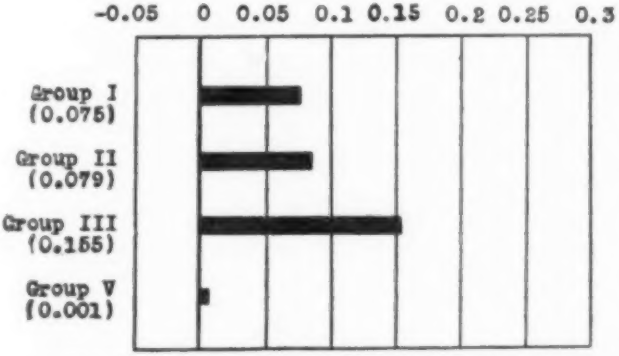

FIGURE 2. LATIN WORDS OF FIRST-YEAR LATIN ORIGIN. SUPERIORITY PER PUPIL PER WORD OF EACH GROUP OVER GROUP IV (ENGLISH CONTROL)—GRADE VII

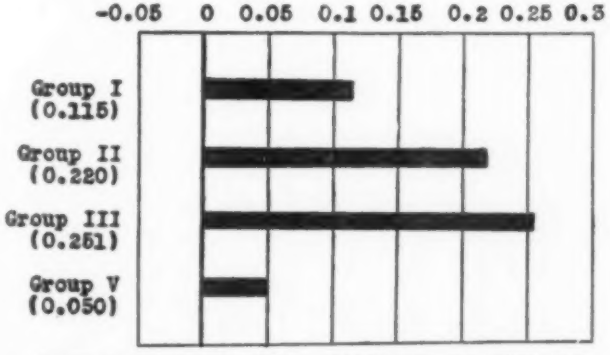

FIGURE 3. PRACTICE WORDS. SUPERIORITY PER PUPIL PER WORD OF EACH GROUP OVER GROUP IV (ENGLISH CONTROL)—GRADE VII

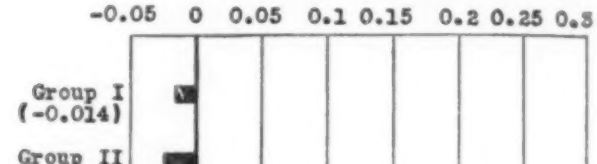

FIGURE 4. NON-LATIN WORDS. SUPERIORITY PER PUPIL PER WORD OF EACH GROUP OVER GROUP IV (ENGLISH CONTROL)—GRADE VII

material. As was true in Figures 1 and 2, the pupils of Group II have a greater advantage over the pupils of Group IV than the pupils of Group I have; and the pupils of Group III have a greater advantage over the pupils of Group IV than the pupils of Group II. The pupils of the non-Latin classes using the special methods show the smallest improvement of all.

Groups I, II, III, and V, when compared in Figure 4 with Group IV, show less excess of gain on non-Latin words, while the pupils of Group II show the least excess of gain on these words. When principles are developed, as in Group III, there is a greater excess of gain than in Group II. The method used in the English Experimental classes (Group V) is most favorable to the spelling of words of this type.

As a result of the study, it is possible to evaluate several methods of teaching as to their effect on the spelling of English words. It is also possible to evaluate several types of words with respect to the help which Latin can contribute to their spelling. As to the methods, we can say that Latin study, even with no special emphasis upon its relation to English spelling, does have an appreciable effect. When the similarities between Latin and English spelling are pointed out, the spelling results are still further increased but the greatest results can be achieved when we develop rules which govern the spelling of Latin derivatives.

Further, this last method produces the least interference with the spelling of non-Latin words.

As to the types of words which Latin study can influence most widely, we can say that there is a slight interference with words of non-Latin origin, that there is a decided gain with words derived from Latin words studied somewhere in the high school, that the gain is still greater with words derived from first-year Latin, and that the greatest gain is produced on words which have been included in the practice material.

The results given above are not changed when we eliminate the influence of intelligence and the personal qualities of the teacher. In a great many cases we were able to control these influences and no appreciable change was found.

THE RELIABILITY OF TESTS REQUIRING ALTERNATIVE RESPONSES

WM. ASKER
Joliet Township High School and Junior College

Several recent tests are so designed that they call for a choice between two or more alternative answers. If in doubt the examinee is urged to guess. The effect of guessing is supposed to be eliminated by subtracting the number of wrong answers, or some fraction thereof, from the number of right answers, the remainder being the score on the test. It is the purpose of this paper to examine the reliability of such a procedure. What are the chances that the score on a test of this kind is a true representation of the knowledge of the examinee?

First, will be examined the chance of a person who knows absolutely nothing about the subject matter of a test to make a passing score through mere guessing. To demonstrate this matter to a group of students unfamiliar with the calculus of probability the following two experiments were carried out.

1. A fictitious *true-false* test of twenty statements was prepared by picking out of a deck of cards ten hearts and ten spades and shuffling these thoroughly. By drawing these cards one at a time the order of true and false statements was determined, hearts representing true and spades false. After the key to the test had been thus prepared, an imaginary individual guessed the answers to the statements, one after the other, by flipping a coin, heads meaning that he guessed the statement was true; tails, that it was false. This same procedure was repeated for twenty individuals giving the following scores: —2, 0, 8, 2, 6, 2, 0, —6, —2, —6, 10, —8, 4, 4, 2, —6, —2, —2, 2, —6. Translated into percent scores the following series was obtained: —10, 0, 40, 10, 30, 10, 0, —30, —10, —30, 50, —40, 20, 20, 10, —30, —10, —10, 10, —30. If 70 or 75 percent is considered as a passing grade, not one individual was able to pass the test by guessing, the highest score being 50. The average of the scores happens to be exactly zero.

2. The second experiment refers to the case where three answers to a question are suggested of which the examinee checks the correct one. The fictitious test consisted of twenty questions with three answers to each. To determine which of the three answers was to be considered as correct seven aces, seven twos, and six threes were picked from two decks of cards and shuffled thoroughly. One of these twenty cards was now drawn. If it was an ace, it meant that to the first question the first answer was to be considered correct; if it was a two, that the second answer was correct; and if it was a three that the third answer was correct. The next card drawn would determine which of the three answers to the second question was correct, and so on. After the key to the test had been thus determined, all the spades of a deck of cards except the ace, the two, and the three were thrown out and the remaining cards shuffled. The number of the first spade turned up determined the order of the answer guessed to be the correct one. After again shuffling the cards a new guess was made in the same way to determine the answer to the second question, and so on until the whole test had been answered. This same procedure was repeated twenty times, representing twenty persons taking the test. Their scores were determined by subtracting one-half of the number of wrong answers from the number of right answers. The following series was obtained: 5, —5.5, —4, —4, 0.5, —1, —2.5, —1, —0.5, 2, 2, —2.5, —4, 2, —2.5, —2.5, 0.5, 3.5, —2.5, 0.5. Translated into percent scores the series becomes: 25, —27.5, —20, —20, 2.5, —5, —12.5, —10, 2.5, 10, 10, —12.5, —20, 10, —12.5, —12.5, 2.5, 17.5, —12.5, 2.5. The highest score is 25, far below the passing mark suggested in connection with our first experiment.

These two experiments indicate that there is a very small chance of making a passing score on either type of test by mere guessing. Of course, they do not prove the matter scientifically, but may be effective in convincing people who are not familiar with calculations of probability, since they may be easily repeated.

Let us now turn to a mathematical analysis of the problem.

If P represents the probability that an event will happen, and Q the probability that it will not happen in one trial, then the terms in the development of $(P+Q)^n$ will, in order, be the probability that it will happen n times, $n-1$ times, etc., down to *no* times in n trials.

In the case of a *true-false* test, there are two possibilities in guessing once. Therefore

$$P = Q = \tfrac{1}{2}.$$

To find the probability of guessing right 20 times, 19 times, etc., in twenty guesses, expand $(\frac{1}{2}+\frac{1}{2})^{20}$ thus:

$$\left(\frac{1}{2}+\frac{1}{2}\right)^{20} = \frac{1}{(2)^{20}} + \frac{20}{(2)^{20}} + \frac{190}{(2)^{20}} + \frac{1140}{(2)^{20}} + \frac{4845}{(2)^{20}} + \ldots\ldots\ldots$$

For convenience in comparing these probabilities the fractions have all been reduced to the same numerator, *one*, and the results brought together in Table I.

TABLE I. THE PROBABILITY OF MAKING VARIOUS SCORES ON A TRUE-FALSE TEST OF TWENTY STATEMENTS BY MERE GUESSING

Right	Wrong	Score		Probability: One in
		Right −Wrong	Percent	
20	0	20	100	1,048,576
19	1	18	90	52,429
18	2	16	80	5,519
17	3	14	70	920
16	4	12	60	216
15	5	10	50	68
14	6	8	40	27
13	7	6	30	13
12	8	4	20	8
11	9	2	10	6
10	10	0	0	5.7

Table I should be read as follows: To guess 20 right, or make a score of 100 percent, there is one chance in 1,048,576; to guess 19 right and 1 wrong, or make a score of 90 percent, there is one chance in 52,429; etc. The negative scores are omitted; the probabilities are the same as for the corresponding positive scores. The probability of making a score of 14 (70 percent) *or better* is one in 776.

In the case of three alternative responses, the probability of guessing the right one in one trial is one-third. Our formula $(P+Q)^{20}$ then becomes

$$\left(\frac{1}{3}+\frac{2}{3}\right)^{20}=\left(\frac{1}{3}\right)^{20}+20.\left(\frac{1}{3}\right)^{19}.\left(\frac{2}{3}\right)+\frac{20\cdot19}{1.2}\left(\frac{1}{3}\right)^{18}.\left(\frac{2}{3}\right)^{2}+\ldots\ldots\ldots$$

which gives Table II.

TABLE II. THE PROBABILITY OF MAKING VARIOUS SCORES ON A THREE-ALTERNATIVE TEST OF TWENTY QUESTIONS BY MERE GUESSING

Right	Wrong	Score		Probability: One in
		Right $-\frac{\text{Wrong}}{2}$	Percent	
20	0	20	100	3,486,784,401
19	1	18.5	92.5	87,169,610
18	2	17	85	4,587,874
17	3	15.5	77.5	282,323
16	4	14	70	44,979
15	5	12.5	62.5	7,028
14	6	11	55	1,406
13	7	9.5	47.5	351
12	8	8	40	108
11	9	6.5	32.5	40
10	10	5	25	18
9	11	3.5	17.5	10
8	12	2	10	6.7
7	13	0.5	2.5	5.5
6	14	−1	−5	5.5

Table II should be read in the same way as Table I. The probability of making a score of 14 (70 percent) *or better* is one in 39,876.

This analysis shows that either of these two tests is practically safe *as far as guessing throughout the test is concerned;* in other words, the probability is very small that a person who knows nothing about the test will make a passing score by guessing. Still safer would be a test with four alternative responses.

But the practical case is different from the one we have been discussing. It would be folly to give a test such that mere

guessing would have to be resorted to. The actual situation will be one where the average student *knows* the answer to some of the questions, and *guesses* at the rest. We shall now analyze the chances that the student's score on the test in this case is a true representation of his actual knowledge. We shall assume that we give a true-false test of 30 statements of which the student actually knows 20, or 66⅔ percent, which would probably be a reasonable passing mark. He guesses at the other 10 statements. To find his chances of guessing the various combinations of right and wrong, we develop the expression $(\frac{1}{2} + \frac{1}{2})^{10}$. The result is shown in Table III which should be read as follows: To guess 10 right and none wrong, giving a total score of 100 percent, there is one chance in 1,024, or it may happen in one-tenth of one percent of the cases.

We see from this table that the actually deserved score of 66⅔ percent is obtained in 24.6 percent of the cases; that in 75.4 percent of the cases the obtained score is at least 6⅔ percent above or below the true score; that in 34.4 percent of the cases it is at least 13⅓ above or below; and so on.

If we take the case of 30 questions with three alternative responses and assume as before that the examinee knows the right answer to 20 of these questions, and guesses at the remaining 10, we must develop $(\frac{1}{3} + \frac{2}{3})^{10}$ in order to find the probabilities of guessing the various combinations of right and wrong answers. The result is presented in Table IV which should be read in the same way as Table III.

Here no score is absolutely correct (66⅔ percent) but 26 percent come within 1⅔ percent of the correct score, and 22.8 percent more within 3⅓ percent. We may, therefore, say that nearly 50 percent are approximately correct, while 82 percent of the scores are not more than 8⅓ percent off. The examinee has a somewhat greater chance (56 percent) to receive a lower score than he deserves, than he has to receive a score that is too high. Taking into account the magnitude of the deviations from the true score, the average grade for a large number of examinees

TABLE III. THE PROBABILITY OF MAKING VARIOUS SCORES ON A TRUE-FALSE TEST OF THIRTY QUESTIONS, IF TWENTY OF THESE ARE KNOWN, AND THE REMAINING TEN ANSWERED BY GUESSING

Guesses		Score		Probability: One in	Percent of Cases
Right	Wrong	Total Right −Wrong	Percent		
10	0	30	100	1,024	0.1
9	1	28	93⅓	102	1.0
8	2	26	86⅔	23	4.4
7	3	24	80	8.5	11.7
6	4	22	73⅓	5	20.5
5	5	20	66⅔	4	24.6
4	6	18	60	5	20.5
3	7	16	53⅓	8.5	11.7
2	8	14	46⅔	23	4.4
1	9	12	40	102	1.0
0	10	10	33⅓	1024	0.1

TABLE IV. THE PROBABILITY OF MAKING VARIOUS SCORES ON A THREE-ALTERNATIVE TEST OF THIRTY QUESTIONS, IF TWENTY OF THESE ARE KNOWN, AND THE REMAINING TEN ANSWERED BY GUESSING

Guesses		Score		Probability: One in	Percent of Cases
Right	Wrong	Total Right $-\frac{\text{Wrong}}{2}$	Percent		
10	0	30	100	59,049	0.002
9	1	28.5	95	2,952	0.03
8	2	27	90	328	0.3
7	3	25.5	85	61.5	1.6
6	4	24	80	17.6	5.7
5	5	22.5	75	7.3	13.7
4	6	21	70	4.4	22.8
3	7	19.5	65	3.8	26.0
2	8	18	60	5.1	19.5
1	9	16.5	55	11.5	8.7
0	10	15	50	57.7	1.7

would, of course, in this type of test as well as in the true-false test be equal to the actually deserved grade.[1]

Our investigation has shown that a true-false test or other tests requiring alternative responses are practically "safe" in as far as guessing throughout the test is concerned; that is a person who is absolutely ignorant about the subject matter of the test has a very slim chance of making a passing score by guessing. But when part of the test is known, and the rest is answered through guessing, there are rather large chances both for undeserving individuals to make a passing grade and for deserving pupils to fail. Especially is this true in the case of a test with only two possible answers.

Tests of this kind possess, however, so pronounced advantages in many respects over the traditional examinations, that they undoubtedly will come into more and more general use, in spite of the attacks that are sometimes made upon them. But a more reliable method of scoring must be found. The effect of guessing *is not* eliminated *in the majority of individual cases* by subtracting the number of wrong answers, or a fraction thereof, from the number of right answers. Such elimination takes place only if we average the scores of a large group, but the purpose of a test is, in most cases, to find out where the *individual* pupil stands. It seems that a more reliable index of the student's actual knowledge would be gained by discouraging instead of urging guessing. It might be a good plan to request the examinee to put a special mark against those questions that he does not know, but determine his score on the basis of right responses only. Such a plan might sometimes pass a student who did not deserve to pass, but there would at least be no risk of failing a conscientious student if he deserves to pass.

[1] In a recent article "A criticism of tests requiring alternative responses," (*Journal of Educational Research*, 6:236-40, October, 1922) H. H. Hahn concludes, on the basis of an experiment performed, that the method of scoring a true-false test would always result in a score higher than the actually deserved one. This conclusion is based upon the erroneous procedure of counting negative scores as zero, even when they refer to that part of the test only, at which the examinee guesses. It may be convenient to call negative scores on the whole test zero, but this must never be done when computing averages.

Editorials

HOMOGENEOUS OR NON-HOMOGENEOUS GROUPING

The purpose of classifying school children is to bring together pupils who are of equal or approximately equal ability in order that common instruction may be given. We have known for a long time, however, that the members of a class organized as classes are usually organized, were not of the same ability; and at the end of each term we have expressed our realization of this fact by giving passing and failing grades to varying proportions of the class. Moreover, we have divided our classes, especially in the elementary grades, into strong and weak sections or fast and slow sections; and we have done this in tacit recognition of the wide range of talent within each class and because we believed there were advantages in a greater degree of homogeneity.

As soon as general intelligence tests became available, a large number of school systems began to use them as instruments for classification, placing pupils of closely similar intelligence quotients or mental ages in the same section. The argument for this has been, of course, that those of like mental ability could be better taught together than those of widely differing ability. Some schools, recognizing that the bright-lazy and the dull-industrious tended to prevent perfect correlation of mental status and scholastic achievement, have used a combination of intelligence-test scores and teachers' judgments of ability as the basis of their grouping. The effort, however, has always been to secure as nearly homogeneous groups as possible.

We were, therefore, somewhat surprised to find in the December number of an educational journal an account of a school

system in which the whole theory of homogeneous grouping had been thrown overboard, with the approval of the teachers and the administration, on the ground that its undesirable results far outweighed its advantages. The writer of this article states that, after the teachers had used homogeneous grouping in departmentalized sixth, seventh, and eighth grades for seven years, on the basis of the previous year's average scholastic standing—few if any intelligence tests having been given—the plan had been abandoned. It was abandoned because it seemed to have but two desirable features, while its undesirable features were said to be nine in number.

The desirable outcomes were: (1) the plan works fairly well with the middle fifty percent of the pupils; and (2) the upper quartile covered somewhat larger amounts of subject matter.

The undesirable features were: (1) the pupils in the upper quartile relied upon versatility and cleverness instead of hard work; (2) this group developed snobbishness and egotism; (3) the pupils in the lowest quartile made little effort; (4) to many the opportunity of getting into higher groups made no appeal; (5) parents did not understand the scheme and nagged the children who were in the lower sections; (6) the teachers favored the stronger and slighted the weaker sections; (7) such classification is undemocratic; (8) since it is impossible to give the same work to all sections, the pupils in the lower sections enter high school handicapped; and (9) intelligence tests show high mental ability among some pupils whose scholastic standing is low.

On the basis of these statements of the teachers, the administration decided to change the plan of classification. Each grade—the sixth, the seventh, and the eighth—had four sections. According to the new plan, each of the four sections of a grade was to have an equal number of pupils from each quartile of the range of scholastic ability. In order to take care of these differing abilities within a section, the teacher was to hold the pupils of the lowest quartile merely for the minimum

essentials and to give to those of greater ability varying assignments involving extra work, reports, etc.

Well, thus we have *change*, whether or not we have *progress*. Let us examine in some detail the alleged undesirable features of ability grouping and the proposed remedy. Points one and two may be considered together. It is by no means clear that placing pupils of the highest quartile in the same class with those who are greatly their inferiors will induce hard work or prevent these bright children from recognizing their own superiority. On the other hand, if capable pupils constitute the entire membership of a class, the conditions appear to be especially favorable for exacting a type of work which requires real effort. Under these circumstances, snobbishness and reliance upon cleverness would tend strongly to disappear. Doubtless there would still be enough range of ability within this upper quartile to make for differentiation. Likewise for point three; in the lowest quartile, some pupils would, of course, be superior to others; and if their work were adjusted to their ability, the joy of accomplishment would appear regardless of how this accomplishment might compare with that of superior groups. At any rate, if minimum essentials failed to appeal to them when they were in a group by themselves, are they likely to do so when these same pupils, in daily and hourly contact with more capable pupils, are made to feel their inferiority?

In regard to the fourth point, lack of enthusiasm for advancement to higher groups, we fail to see how the new plan will improve matters. This particular form of motivation failed, but what has the new plan to offer in its place? Numbers five and six are both administrative problems. Parents who nag their children for being in low sections are quite likely to nag their children if they receive low marks. Will the new plan provide high marks for those who do the best of which they are capable even though it be only the minimum essentials? And will the plan of having all degrees of ability in the same class safeguard the weak pupils against those teachers who formerly slighted them and favored the strong?

The seventh point is most serious. Undemocratic! What giants have been slain and what policies cast into outer darkness by the invoking of that potent word. It will scarcely be thought undemocratic to assert that children are unequal in ability. The fact of such inequality is generally admitted. Moreover, both of the plans which we have been considering are based upon the assumption of inequality. The undemocratic part seems, therefore, to lie in the grouping. But to seat pupils in class in such a manner that those who are mostly alike may work happily upon a common problem within their comprehension; and to seat together another group so they too may succeed by conscientious effort on their own level is not undemocratic. We wonder if, in the school system to which we are referring, the trouble was not really one of curricular adjustment and proper method of instruction.

The eighth point asserts, as a defect of the grouping plan, that pupils of the lower sections entered high school handicapped. We have no reason to believe, however, that those who formerly belonged to the lower groups will now, under the new plan, enter high school with any better equipment. Unless they do, the handicap will remain. But more important still, why should all pupils, regardless of ability, be given the same work? Shall we never learn that teaching can be measured only by how much our pupils really learn and put to use and not by how much subject matter has been assigned and has been gone over in class?

The ninth point, that high intelligence is possessed by some pupils of low scholastic standing, is not an argument against classification as such. It may mean that all the instruments available for classification have not been adequately used; although it is admitted that even under the most favorable circumstances, grouping will still fall short of complete homogeneity. Some discrepancies will still remain. This points to the fundamental fact that the problem of doing effective school work is not solved by classification, even when classification is

made as accurately as possible. Unless we modify our method, adjust our curricular offerings in terms of the ability of these different groups, and administer the whole scheme wisely, we might as well not classify at all. Classification is but one means to the end of caring for individual differences in children and it profiteth little if all the other means to this same end are disregarded.

E. J. A.

Reviews and Abstracts

STRAYER, GEORGE D., AND ENGELHARDT, N. L. *School Records and Reports.* New York, Bureau of Publication, Teachers College, Columbia University, 1923, 81 pp.

The city superintendent, in recent years, has found himself between the Scylla of keeping records and the Charybdis of making reports. On the one hand he has faced the growing impatience of the teacher who feels that so much of her time is wasted on the mechanics of her job that her real task is neglected. On the other hand, the survey commission has caused him to realize that if he is to justify the expenditures which he is making for education, he must have facts regarding his school system which he has often not been able to supply; and moreover, that these facts must be properly reported.

This book is a first step in giving superintendents the necessary directions for establishing and for judging a system of school records and reports for a city system of 100,000 and less. The score card which is found on pages 9-12, indicates the method used in determining the efficiency of the city's system of records and reports. One thousand points on this card indicates a perfect system. Nothing, however, has been suggested indicating what a score of 300, 600, or 900 points might mean. There are 181 items (records, reports, or other forms) required. All of these are not required in order to have a 1,000 point score, however, for the authors suggest "in case a record is proven a non-essential in a school system full credit may be given." The greater part of the book is given over to a discussion of each of these 181 items. Standards are set up indicating what the composition of each must be if credit is to be given on the score card for such an item.

The entire report is, according to the authors, only a preliminary edition and they are hoping to receive suggestions and contributions that will help them in making revisions. Pages 73-79 are left blank; and on these pages a criticism of the general plan together with proposed changes in any of the subheads of the score card is asked for. Even in its present form, however, the card cannot but be very helpful to the progressive city superintendent.

ARCH O. HECK

Ohio State University

CADY, VERNON M. *The Estimation of Juvenile Incorrigibility.* Whittier, California, Whittier State School, 1923. 140 pp.

This investigation has as its purpose an attempt "to find to what extent it is possible, by means of the test method, supplemented by character ratings, observational data, and other aids, to identify, in advance of overt delinquency,

children of abnormal moral tendency." It is an ambitious, though rather inadequate effort to apply direct measures to a problem that has so far foiled all theoretical as well as experimental attacks.

The author shows no orientation in the general field of delinquency and ignores such theoretical analyses of his problem as are available in the literature of psychology and criminology. There is an almost complete lack of reference to earlier work on the specific problem. His own approach is a direct frontal attack by means of tests, questionnaires, and rating scales, unsupported by any serious attempt at psychological analysis.

A closer examination of the experiment is hardly more encouraging than the general view. The lack of perspective and orientation is not made good by clearness of original ideas. Four divisions of "personality" are proposed, viz: "(a) intelligence, the primary means of making objective adjustments; (b) energy-drive relations, the capacity for evoking sustained internal stimuli; (c) preferences, directions, tendencies, the seat of morality, group selections, worth relations; (d) temperament; an equivocal term expressing manner of response and probably greatly influencing c, b, and a, in this order."

Two thousand and twenty-four boys, aged 9 to 19, were tested with a "moral test battery" of 5 tests in groups of about 25 in 20 California schools. Seven hundred teachers rated the children. However, only 150 boys, "very corrigible or very incorrigible," aged 12.5 to 14.5 were included in the final study. Following an elaborate statistical analysis the author concludes, that "social adjustment may be measured with a fair degree of certainty," that "so-called moral qualities are probably distributed in the population more in the form of a normal curve than has been supposed," that "it is possible that moral disposition is relatively fixed in the individual, though modifiable by the growth of intelligence and the accumulation of experience and habit," etc.

Maybe so.

EDGAR A. DOLL

Ohio State University

PINTNER, RUDOLPH. *Intelligence Testing.* New York, Henry Holt, 1923. 406 pp.

This book is a simple account of the history, theory, technique, and results of intelligence testing in this country and abroad, designed as a college text and as a guide to teachers. The book is complete in the scope of the subject matter, which makes it very valuable as an outline or guide for college students. The author shows an excellent grasp of the whole field of intelligence testing and a commendable desire to give some treatment to all phases of the movement. The comprehensive bibliographies which follow each chapter, are themselves witness of the range covered. The style is dignified yet simple, and the treatment complete though brief.

Beginning with the time of Greece and Rome the author first traces the general history of intelligence tests with relation to social, psychological, and anthropological trends. The work of Binet receives an entire chapter in which is given one of the few adequate and accurate accounts of those important

developments. In this and in the description of the advances since Binet, the author gives credit where due and reveals his own intimate personal contact with the movement.

In a particularly valuable chapter on the concepts of intelligence the author succeeds in presenting the main issues simply and without controversial discussion. One regrets the lack of space which prevented a more detailed treatment of this important material which is all too often omitted in other works on mental tests.

The book is in three parts: (1) historical and theoretical, (2) methods, (3) results. One wishes that each part might have been expanded in a separate volume. The part on methods describes the most commonly used group tests and the scales used for individual testing. The latter includes all three original Binet scales as well as the principal American revisions. The account is descriptive and avoids details of technique. No account is given of single tests not in scaled arrangement. A comprehensive account is given of group tests (not including strictly educational tests) which includes the most recent developments.

The chapter on results is an important compilation of the more important experimental data obtained by applying tests to feeble-minded, superior, delinquent, dependent, deaf, and blind children as well as to soldiers, employees, college students, and school children in general. There is also a chapter on tests of negroes and another on tests of the foreign-born. The reader will wish to consult these for details and will be gratified at finding comprehensive treatment and extended bibliographies.

The book as a whole will prove most useful, especially if supplemented by the detailed references to sources. We fear, however, that unless so used the volume by itself is somewhat too sketchy for college students; while at the same time it is likely to be too detailed, that is, lacking in generalization, for teachers.

EDGAR A. DOLL

Ohio State University

BRIGHAM, CARL C. *Study of American Intelligence.* Princeton, Princeton University Press, 1923. 210 pp.

Much has been said and printed about the results of psychological examining in the Army during the recent War. The complete results of that important undertaking have been published in Memoir XV of the National Academy of Sciences. But the material is so voluminous and so very complicated that none but the technical expert can read it with profit. It is indeed, rather a source book of methods, data, and statistical results of the examinations of more than a million and a half recruits—the first attempt of any nation to scientifically classify its military man-power.

That the compilation of material in Memoir XV is of the utmost significance for our national welfare, is well recognized by all students of politics, education, and the social sciences who have been able to cope successfully with the unusually complicated nature of the material. One must be psychologist, statistician, educator, anthropologist, and sociologist, to grasp the meaning of

the wealth of data presented and analyze the intricately involved relationships. The real significance of psychological examining in the United States Army will therefore not become available to any large body of the general public or even to the scientific world until the original data have been resolved in a succession of studies dealing with its most important indications.

Such a study, the first adequate treatment of a particular phase of the material in Memoir XV, is presented in this volume. Mr. Brigham is peculiarly well fitted for his arduous undertaking by reason of his own war record, first in rehabilitation hospitals in Canada, then in our own ranks as a soldier, then as one of the first examining officers when psychological examining of recruits was in its earliest experimental stages, then as officer in the directing and analyzing headquarters of the Division of Psychology in the Surgeon-General's office, later as adjutant in the Tank Corps, and finally as director of vocational rehabilitation after the armistice. This unusual service gave Mr. Brigham first-hand acquaintance with the material and problem with which he deals that adds much to the practical value of his conclusions. He is also well qualified by reason of his broad academic and scientific preparation, not only in psychology, but in the fields touched by the material studied. Let the critical or skeptical reader note, then, not only the obviously classical treatment of data, minute and comprehensive logical analysis, painstaking statistical labor, and precise statements of results, but also the personal and professional perspective of the author.

A critical review of Mr. Brigham's study is hardly possible without a more detailed analysis than is here desirable. Moreover, such specific objections as might be made against its standpoints, argument, and conclusions are neither important nor becoming, in contrast with the outstanding merits of the study as a whole. We do not entirely agree with his somewhat dogmatic statements regarding the reliability and validity of the group examining procedures, though to be sure his space prevented a detailed justification thereof. We regret that the individual examining procedures were not evaluated in more critical detail. We cordially disagree with the disparaging attitude toward the "mental age" concept, which is like striking the hand that feeds and is also theoretically weak. We find inconsistencies in the details of arriving at the conclusions regarding the relation of foreign-born intelligence to years in United States. We think the section on the "race hypothesis" reveals internal inconsistencies, is not essential to the main argument, and is none too convincing, though we admire the author's endeavor to meet this most baffling problem. But however we may fail to follow his argument in detail we are in close agreement with his major conclusions in our own analysis of the relevant material. Moreover, the study has one admirable feature which offsets all minor failings, namely, the thoroughness and honesty of treatment.

The book begins with a concise description of the examining procedures employed by the army psychologists, with a brief evaluation of their adequacy. This is followed by a description of the sampling used as the basis for Memoir XV and the present study. The principal purpose of the study is an analysis of the intelligence of the foreign-born recruits as a whole, by separate nationalities, and in comparison with negroes, native white recruits, and white

officers. It is found that the intelligence measures of the principal foreign-born groups are decidedly below those of native whites. The unit of measurement is the "combined scale," a device for pooling the results of the several types of army examining procedures. In particular it is found that the intelligence of Negro recruits, Italian recruits, and Polish recruits are decidedly inferior to other important groups and to the native white recruits.

The main study of the intelligence of the foreign-born is specifically elaborated by an analysis of the findings in relation to length of residence in this country. This reveals that the intelligence measures decrease steadily with recency of residence, especially for those recruits who have been here less than 15 years. The further analysis of this important result is the most difficult and intricate part of the study and yields some of its most important implications. This section is difficult to follow, though clearly set forth, and we fear that the biased reader may all too readily refuse to subordinate emotional feeling to intrinsically difficult analysis.

In an attempt to generalize on the nationality data the author employs a race hypothesis that reduces the nationality groups to three racial groups, Nordic, Alpine, and Mediterranean. This is debatable from many angles, but the author confirms his earlier findings in more general terms.

These results as well as those of the study as a whole are then examined with reference to the problem of immigration in this country. It is here that Mr. Brigham's study transcends psychology and scientific treatment in the import which it has for our national welfare, immediately and ultimately. It is here that his study indicates the grave need for such dispassionate investigations if we would maintain our national integrity and superiority.

The book is decidedly well written and is readable in spite of the intricate and technical character of the analysis. The author acknowledges his debt to Memoir XV and to those psychologists, Boring, Brown, and May (also mentioned in the foreword by Col. Yerkes) whose statistical resourcefulness made that work possible.

EDGAR A. DOLL

Ohio State University

WASHBURNE, CARLETON W. *Individual Speller.* Yonkers, World Book Company, 1924. 76 pp.

This book presumes to present to pupils in grades III to VI, inclusive, about 3,600 words which are most important for children to study in a spelling text. The word list is made up from the W. N. Anderson study of adult correspondence, the Thorndike study of reading vocabulary, and the Tidyman study of children's compositions. The assumption is that words of high frequency in children's writing, in literary reading, and in adult correspondence should be combined to determine the words which children should learn to spell. Many investigators have seriously questioned whether a combination on these three bases should be used; but the author has frankly stated the sources of his material in an article published in the *Elementary School Journal* for June, 1923.

The words are arranged alphabetically by semester lists, and columns are provided in which checks may be made to indicate misspellings, an O. K. being placed by the teacher when the child has learned the word. A method of study is given very briefly in the Directions which tell the pupil how to study and how to administer the work with his partner. Dictation exercises are provided at the end of each semester list, but children are asked to write the significant words only. In addition to the word lists and provision for checking, the book contains nineteen ruled pages upon which children are expected to write words which are misspelled in their written work.

By presenting the words in alphabetical order, ease of locating words is increased, but the range of difficulty within any one lesson upon which a child is working is considerably increased. The method would undoubtedly have been improved by the provision for some review between the end of the week in which the word is first studied and the beginning of the fall semester.

E. J. ASHBAUGH

Ohio State University

GATES, ARTHUR I. *Psychology for Students of Education.* New York, The Macmillan Company, 1923. 489 pp.

What psychology is of most worth to teachers and how it shall be presented are matters upon which there is little agreement. Educational psychology is the course in education which is most frequently specified in the requirements for the certification of teachers. As ordinarily given, it involves a prerequisite, of course, in general psychology. This book, admirable as it is in many ways, does not quite fit the situation either for general psychology or for educational psychology. It contains too little educational psychology and too much general psychology for a course requiring the latter as a prerequisite, and, on the other hand, has too little general psychology and too much educational psychology for the ordinary introductory course. It is, however, better adapted for the latter use. The author indicates in the Preface that he has had in mind its possible use as a text in general psychology with an educational flavor for he says, "While the book has been written to satisfy the needs of students of education primarily, it may be found serviceable as an introductory text in general psychology, especially by those who wish to give a relatively large amount of time to the dynamics of human behavior and the more complex activities." The label "for students of education" will, however, probably prevent its use as a text in classes for which it is best adapted. The book is scarcely more specifically a textbook in educational psychology than, for example, Woodworth's *Psychology*, which it much resembles. But, as matters stand now, the label will probably militate against its use as a general psychology text.

For institutions and classes where general psychology and educational psychology are combined in a single introductory course, the book is an altogether admirable text. One hundred pages are given over to the physiological substratum of human behavior and to conscious processes, another hundred to man's native equipment in instincts, impulses, and emotions; two hundred

to the psychology of learning, the transfer of training, and the effect of fatigue and drugs on efficiency; and almost another hundred to individual differences, correlations, and the measurement of intelligence. This gives a rough idea of the balance of the book and the relative emphasis given to the different topics treated. It is clearly written, the topics are generally selected with care and discrimination, and the questions and exercises at the close of each chapter are much better than usual.

It may be that so-called general psychology as a prerequisite to other special brands will "go by the boards," in which case this book, and others of a similar mould in other fields than education, will find a wide use. For a combined general and educational psychology course, the author has provided an unusually fine text.

V. A. C. HENMON

University of Wisconsin

MEAD, ARTHUR RAYMOND. *Learning and Teaching.* Philadelphia, J. B. Lippincott Company, 1923. 277 pp.

This book devotes approximately 134 pages to exercises, both review and advance, experiments to be worked by the student, and references—which latter are valuable for their specificity. The justification for such an abundance of "additional" material—and at first sight it might indeed seem an overabundance—is given by the author in his Preface, "The material presented in this volume and the methods suggested for its use have as their chief aim the identification of theory with practice, or the application of information to actual problems of the school" (Page vii). It is Doctor Mead's intention that the courses in education and psychology in which his book is used as a text shall bear "obvious, direct, and useful relationship to the teacher's work" (Page vii). While one would not recommend this as a book to be read for its original contributions, one could give it hearty support as a text in educational psychology.

In a sense, the title gives away the plot. "It is the plan of the book to make the relations between learning and influencing factors in learning the constant line of study" (Page 2). This is, in brief, a bird's-eye-view of the direction the work takes. If now, we but understand the manner in which learning is conceived, the development step by step must be, on the face of things, evident enough.

The author acknowledges a great "debt to Doctor E. L. Thorndike" and we are not surprised, therefore, when we are informed that in a simple learning situation the three elements are: "S–R, in which S is stimulus (stimuli, situation), R is the response (reaction) made to S, and — is the connection (bond) between the two" (Page 6). Nor are we surprised when we read "(1) that one learns by doing, or using responses, and (2) that the learning is perfected by repeating the responses" (Page 6). By the time we have finished Chapter II we have had illustrated for us the following laws of learning: "Use, Frequency, Disuse, Effect (including both satisfaction and annoyance), Multiple Response, Partial Activity, Influence of 'Set' or Attitude,

and Analogy" (Page 8). Among the influencing factors in learning we find five types of interests that "have been designated by Thorndike" (Page 62); and his "ease series" is given as "three other conditions" which are "favorable factors in learning" (Page 66).

This is all by way of supporting our previous statement that we find nothing fundamentally new in the book. We do find, however, an organization of material which forms an "ease series" all its own for the future teacher. The book proceeds from simple learning conditions up to and through aims of education. Each chapter starts with a statement, in bold-face type, of the points to be developed. These are then enlarged upon and the chapter ends with a summary of the presentation—the main points again being stressed. Exercises, experiments, and references follow. This is all in line, of course, with the formation of the desired S–R series. We may say, therefore, that the book has two outstanding traits for which it is to be commended—namely, its organization and the opportunity it affords for range and detail in study.

The chapter on aims of education bulks in no wise heavy, which, when one considers the author's statement that "high-grade teaching" is "the central theme of this book" (Page 112), seems an oversight; and, if deliberate, unfortunate. Aims are discussed both sociologically and psychologically, and one is tempted to ask, why not from the point of view from which they must ultimately be decided—namely, from the point of view of philosophy?"

We find the author saying, "Just *what* habits to insist must be formed, just *what* ideals to secure, just *what* information to teach children, just *what* interests and desires to arouse and cultivate, is not easy to ascertain" (Page 161). It is one thing to say "Teaching is the establishment of S–R, and the aims of teaching in relation to S–R are to establish R and connections between R and S" (Pages 165-166), but is it not a horse of another color to ask "What S to present" and "What R to seek?" Since the "what" is left unanswered, this seems but an admission that the sociological division supplies no answer. The sociologist may determine the "what"—the objectives—that have been of value in promoting social growth, but if he never gets off this level then his business is nothing more than the work of maintaining the *status quo*. If he does depart from this scheme to ask "what objectives should we follow," rather than "what objectives have we followed" then he is no longer just a sociologist. It is at this point that he makes use of his scheme of life—his philosophy—and he evaluates his objectives on this basis. If the author does this, then he is misleading when he credits sociology alone.

One cannot help expressing some surprise at the statement that Chapter IV, "has emphasized two types of learning in which analysis and selection are predominant, *thinking* and the *acquisitions of abstractions*" (Page 42, italics mine). Such a division seems, on the face of it to give impetus to a mechanical routine which is hardly in keeping with the work of one who accepts, as the author does, Dewey's discussion of thinking.

H. Gordon Hulfish

Ohio State University

News Items and Communications

This department will contain news items regarding research workers and their activities. It will also serve as a clearing house for more formal communications on similar topics, preferably of not more than five hundred words. These communications will be printed over the signatures of the authors. Address all correspondence concerning this department to Doctor E. J. Ashbaugh, Ohio State University, Columbus, Ohio.

There were 137,147 essays and 3,732 posters on alcohol and tobacco entered in 1922-23 by public school pupils in the contests annually conducted by the Woman's Christian Temperance Union.

Brookline, Massachusetts, has a curfew at five o'clock, designed to call the children off the streets and remind them of home lessons to be done. Town officials hope that this will not only be beneficial to the school work, but will very much decrease the danger of accidents.

The report of a subcommittee on the junior high-school program study of the National Council of Education says that on the basis of a vote among teachers and administrators, 75 percent are in favor of at least a six-hour day, 60 percent are in favor of sixty-minute periods, and 100 percent are in favor of supervised study.

Oklahoma is putting on an intensive campaign this year to eradicate illiteracy. Much should be accomplished by this effort, since the press, the schools, the club women, the American Legion, and many other organizations and individuals are actively cooperating in the attempt to solve the problem.

A recent bulletin from the Bureau of Research of the Minneapolis schools gives the result of a survey of reading, arithmetic, and spelling in schools organized on the departmental and junior-high bases. Slightly greater efficiency, measured by the tests, is shown by the departmental grades in contrast to the junior-high organization.

The School of Music of the University of Wisconsin has announced an all-Wisconsin high-school music contest at Madison for May 15 and 16. Contest events consist of girls' glee club, mixed chorus, orchestra (classes A and B), band (classes A and B), solo (girls' voices and boys' changed voices), piano, and violin.

The election of William McAndrew as superintendent of schools of Chicago is a matter of national interest. The second largest city in the United States has had its full share of educational troubles, and everyone interested in education will extend to both the city and the new superintendent the sincerest wishes for real success in educational development.

A new law in Latvia requires instruction on the dangers of alcohol to be emphasized as a part of the training in health. The instruction must be given in all private schools, and in all public primary and secondary schools. The minister of public instruction must revise the school program within a year to include the instruction.

What is your P.Q.? It is quite as important for teachers and pupils to have a high P.Q. as it is advantageous for them to have a high I.Q. When you see this symbol in educational literature or hear it mentioned in an address, do not look ignorant, but recognize it as the short name for physical quotient. The term is in regular use by physical-training authorities, and is based on a large number of physical-efficiency tests.

A group of studies in public school finance, prepared under the direction of Professor Fletcher Harper Swift, is being published by the University of Minnesota in its series of research publications. Volume II, entitled *Public School Finance in Massachusetts, New York, and New Jersey*, is now ready for distribution. The price of this volume is two dollars. Orders should be addressed to the librarian of the University of Minnesota.

Professor W. A. McCall of Teachers College, Columbia University, who has spent a year in China working with the leading professors of education and psychology in that country, reports that under his direction ten different intelligence tests, thirty different school achievement tests, a special non-language intelligence test, and a point system for standard athletic games were prepared, scaled, and standardized. The tests were built on American models; but, in most cases, he believes they are superior to the originals.

The National Foreign Trade Council, 1 Hanover Square, New York City, has recently published a thirty-page bulletin on *Our Imports and Who Use Them*. The pamphlet is intended for teachers of geography, economics, and foreign trade in grammar schools, high schools, and universities. It is well illustrated and presents in usable form a type of material not often readily accessible to teachers of these subjects. The bulletin can be secured for ten cents from the address given.

The annual reports of the chief examiner and director of research for the United States Civil Service Commission is of especial interest to research workers. The division of research has devoted a large amount of time and energy to devising tests for various types of clerical work. Tests to replace the requirement of high-school graduation and others to determine aptitude for

various types of clerical service have been devised and standardized. Sufficient items are included to permit the making of a number of forms of equal value.

The people of Long Beach, California, proved their support of education and faith in the program of their school officials in January by endorsing a bond issue for $4,900,000 by a vote of approximately 19 to 1. The building program which is to be cared for by this issue will involve the enlargement of the present accommodations and the erection of new buildings for elementary, junior high, and senior high schools. The issue had the united support of the Parent-Teacher Association, Chamber of Commerce and other civic bodies, and organized labor with its women auxiliaries.

The January number of the *Journal of the Florida Education Association* contains a score card for rating teachers devised by a graduate class at Peabody College. Twenty-four score cards were examined in making up the score card presented. It is of interest to note that discipline received the highest point score; care of light, heat, and ventilation, second; and the results obtained in teaching, third. The least important points judged by the point score given are ability to meet and interest parents, adaptability and resourcefulness, and enthusiasm.

The various organizations interested in the teaching of citizenship will find much material in the federal census not only to justify their activity but to urge them to additional endeavor. With more than five million foreign-born children in the schools of the nation in 1920, and with the population of four of our eastern states ranging from 66 to 70 percent foreign-born or native-born of foreign-born or mixed parentage, it is evident that much work must be done if American ideals and traditions are to be fostered and respect for law and order maintained.

The *Baltimore Bulletin of Education* for December reports in part the results of the "Springfield tests of 1846" which were given to the children of the city schools during the past year. The Baltimore children obtained ten percent better ranking in spelling—even on the type of words used in 1846—than did the pupils of that day. They were nearly fifty percent better in arithmetic, although the test contained some problems which would not be used in present-day teaching.

It is the opinion of District Superintendent J. S. Tildsley of New York City that for the present at least the departments of French and Spanish have reached their peaks in pupil-load. For this reason he recently recommended to the board that no additional teachers in these languages be appointed. One of the factors which has made for cessation in the development of these departments is a recent ruling to the effect that students rated among the lowest twenty percent in scholastic work at the time of entering high school are not permitted to take work in mathematics or foreign languages.

Score cards for rating teachers have been produced in considerable numbers and the idea discussed in educational meetings and periodicals for some years. Rating cards for high-school principals and other administrative officers have more recently made their appearance. The December 27 issue of *Moderator-Topics* goes a step further and presents tentative standards for rating rural school-board members. We heartily endorse the portion of the discussion accompanying the card which says: "A board member who would score 'yes' upon every question would be an invaluable asset to any community."

Under a law passed by the last legislature in Ohio, students entering any of the teacher-training institutions, whether county or state, must take a battery of tests prescribed by the State Department. Included in the battery this year was the Terman General Intelligence Test and the Otis Self-Administered Test. It is of interest to know that approximately the same number of students took the examination for entrance to county normals as to state and private institutions; and that the median scores on the two groups were almost identical. This is the best answer that has yet been found to the generally expressed opinion that county normals attract a poorer grade of students than state institutions.

Superintendent John C. Stiers of Washington, Pennsylvania, sent us recently a copy of the mimeographed letter which he presented to the members of his school board. Superintendent Stiers is urging the adoption of a building policy. He has mobilized the facts regarding present conditions and probable future needs, and has made an estimate of the cost through the next twenty years of such a bond issue as he believes necessary in order to start the program properly. We shall look forward with interest to a further communication from Superintendent Stiers regarding the action of the board and the community. Such mobilization of data can hardly fail to be effective when presented to the thoughtful people in a community.

The Virginia State Board of Education has announced the following legislative program:

1. An increase in the common school fund of $1,500,000 to encourage a standard nine-months term and a well-trained teacher for every rural school in the State.
2. An equalization fund of $500,000 to be distributed among the weaker divisions of the State to induce uniform educational opportunity.
3. An increase in the high-school fund of $300,000 so that the State can carry a larger share of the maintenance of high schools.
4. An increase of the State Library Fund from $3,000 to $10,000.

State Superintendent John Callahan of Wisconsin sets forth the following educational program for that state:

1. A state public school fund of ten million dollars so distributed that each community can furnish, with this state aid, adequate educational facilities for its children.
2. A teacher with a minimum academic and professional education of two years beyond high-school graduation for every elementary classroom.

3. A teacher with a minimum of four years of academic and professional training obtained in a normal school, college, or university, for every high-school classroom.
4. A good high school within the reach of every boy and girl in the state.

The California League of Women Voters has issued a helpful pamphlet entitled *A Study of California's Method of School Financing*. This booklet, which may be obtained from the League, 233 Post Street, San Francisco, at a price of twenty-five cents, is divided into eight lessons. These lessons have to do with the questions of taxation; the general plan of school finance in California; the problem of increasing school costs; state, county, and district aid; teacher training; and a listing of some special school problems.

Characterized by clearness and simplicity of statement, the facts in this pamphlet are gathered from the best sources and present every evidence of being reliable. One cannot help thinking that a pamphlet of this sort might be well gotten out in every state of the Union. Most states need it more than California does anyway.

A meeting of the British Association for the Advancement of Science will be held at Toronto, from August 6th to 13th, 1924.

The Association has recently added to its various Sections a Section of Psychology. The Organizing Committee of the Psychological Section believes that the meeting will afford an unique opportunity for intercourse between psychologists of Canada, the United States, and Great Britain; and accordingly issues an invitation to any psychologist in the United States or Canada, who may desire to do so, to join the Association as a member for the Meeting, and to contribute a paper or attend the discussions.

Papers should not require more than 20-30 minutes to read; and intending contributors should send a notice of their papers before the end of March to L. Wynn Jones, the Recorder of the Psychological Section (British Association), The Psychological Laboratory, The University, Leeds, England. Further particulars as to membership for the meeting can be had from the Secretary, British Association, Burlington House, Piccadilly, London, W. 1., England; and notices of the general arrangements will be issued in due course by him and by the Honorable Local Secretary in Toronto.

In the November number of the *Journal of Educational Research*, Mr. Garry Myers gives some examples of erroneous ideas obtained by children from their reading. Such an excellent illustration occurred in connection with the article itself, I cannot refrain from describing it.

The first sentence of the article says "Peggoty of *David Copperfield* thought hides a kind of tropical fruit." I read the sentence and my impression was that *thought* hides something. Reflection led me to inquire why should *thought* be hiding any thing, and what sense was there in the statement? My first impression was that *thought* was the subject of the verb *hides*.

Now I submit that I am a reasonably intelligent man. I can score somewhere from 45 to 74 on the Army Alpha test, and could reach the grade of C. Such being the case, some one else than myself is responsible for the incorrect

first impression that I got. Although I have read *David Copperfield*, I am not a lover of Dickens, and could not immediately have located Peggoty. The reason for that is another story. But why could not my friend Myers have made use of the resources of the English language and have said that "Peggoty thought that hides were a kind of tropical fruit," thus avoiding the ferociously difficult construction of the attributive object? What excuse also can he make for separating the subject from its verb by the distance here employed, and not even setting off the parenthetical expression that intervenes by commas?

I would like to suggest that the science of psychology needs to be described in terms that can be understood, and that a fairly large proportion of psychological books and articles appear to be written by persons who did not wish their readers to understand them. Misconceptions are unavoidable, and a new Nobel prize ought to be offered to the psychologist who will write a book in such language that is can be understood with the least difficulty and with the fewest opportunities for misconceptions.

NATHAN A. HARVEY

State Normal School, Ypsilanti, Michigan

Doctor Charters tells an interesting curriculum story. At the Buffalo meeting of the Committee on Apprenticeship Training of the American Construction Council early in December the question of training apprentices by means of a course of study was under consideration. The construction trades are in need of recruits, and the time therefore appears ripe for facilitating the training of apprentices. In this connection the job of curriculum making undertaken by the American Tile Manufacturers Association is reported. Since news of this sort seldom reaches educational people, this note is offered.

Mr. L. S. Hawkins was employed to build a curriculum for use in teaching tile setting.

The first step was to divide the job of tile setting into specific activities. This resulted in separating nineteen fundamental tasks and twenty special tasks, the latter having reference to particular kinds of tile setting. The next procedure was to employ an expert tile setter who carefully described to a "student" the performance of the first of the nineteen activities. The student wrote down the tile setter's directions and then brought them to him for verification and criticism.

From this point on the procedure was directed toward the perfection of the first lesson—i.e., the directions for the performance of the first activity. The student first put the lesson to the test by trying to perform the task with no other guidance than his notes. Whenever he got into difficulties the lesson was amended to take care of the trouble.

But this was not all. Twenty young men were employed for three months to learn to set tiles. Meanwhile all of the nineteen lessons corresponding to the fundamental activities had been completed to the point indicated for the first lesson. These lessons were used as the basis of the work, and wherever difficulties were encountered the written material was revised.

The result was a tile-setting curriculum of unusual merit. It is significant, however, that it cost $30,000 to put over this project. There are several facts of interest to educational people in this account. Not the least of these is that real curriculum construction costs money—much more money than superintendents or boards of education realize.

Supplementing our editorial comment in the January issue of the *Journal of Educational Research*, Director Ernest C. Witham of the research department of the public schools of Wilmington, Delaware, had this to say:

"Your 'Annual and Semiannual Promotion' editorial interests me, because I have never believed the semiannual plan a good one. In addition to the strictures you put upon the system, I would like to add three others of the first magnitude which you do not mention.

1. *A more frequent change of teachers.*—Under the yearly promotion plan it is common practice to have a teacher remain with a class one school year. In most systems where they have the semiannual promotion plan, teachers remain with a class only half a school year. Of course, with either system it is possible for the teacher to be promoted with the class, but this is uncommon, and hard from a practical standpoint to administer. Only on the assumption that a teacher is very poor is there argument for changing her pupils more often than annually.

2. *A great increase in the administrative work.*—It takes two weeks, at least, and usually a month to settle up one half-year and start another. The size of the classes entering or starting in September are always larger than those starting in February. This involves careful planning, moving of school desks from room to room, and many other changes necessary to make the school run smoothly. The pupils just about get settled down to doing good work when, under the semiannual plan, the arrangement has to be worked out all over again.

3. *Doubles the anxiety periods of the pupils.*—Every June we read in the paper of some tragedy where a pupil has failed or thought he was going to fail, and rather than be disgraced, as he sees it, in the eyes of his peers, he ends it all. It is not only on account of the few poor victims such as this that promotion periods are hazardous. At such times, conscientious teachers are over-worked and do very little real teaching; and many pupils are wrought up to a very nervous state which is unhygienic and frequently very damaging.

4. *Bad adjustment with higher educational institutions.*—Pupils who graduate from the high school in January and who desire to go to normal schools and colleges have a half-year on their hands which oftentimes they are unable to dispose of to the best advantage because normal schools and colleges are run on the annual plan and admit classes only in September.

Semiannual promotion does not in practice make for a more flexible grading system since it has nothing to do with abilities to do school work. In fact there is only one way really to work out a system of flexible grading and that is, to divide pupils into groups according to their ability to do school work and then to promote them on three or four different planes of advancement. This means that the course of study must be arranged for the three

or four different groups of ability within each grade. The 'C' and 'D' groups would be confined largely to minimum essentials and drill work, while the 'A' group would need but little drill, much library work, and a very full and complete course of study. The middle group would pursue a course midway between the other groups. At any time pupils should be transferred from one group to another or from one grade to another, if such change better adapts the school to their needs.

This system does away with the pernicious fallacy of trying to educate anyone by subjecting him to impossible tasks, his abilities taken into account; and also that other pernicious fallacy of trying to educate plastic human beings, —boys and girls—by a process of systematic failure. The time has long past when it is necessary to maintain the high standing of a school at the expense of flunking the weaker member of the class. A high standing never was so easy to obtain as that."

Reading Tests Covering Entire Books

The criticism is frequently made that most of the standard tests in silent reading necessitate a type of reading process not typical of reading as ordinarily carried on by children or, especially, by adults. The usual reading test consists of a series of scrappy paragraphs, having no relation to each other, each paragraph being followed by one or more questions. Presumably a major element in successful silent reading is ability to get the larger relations between masses of facts, as presented in a lengthy coherent treatment of some subject. This type of ability is clearly *not* involved in any of the reading tests at present available.

In an effort to get away from this difficulty, Doctor L. C. Pressey has recently developed a series of reading tests covering certain supplementary readers for the lower grades.[1] Each test covers four or five chapters of a book and consists of a four-page folder having directions on the front page, formal test questions on the middle two pages (the children underline the correct answer, of four answers offered for each question), and thought questions, to which the child writes the answers, on the back page. The tests thus cover the entire book in large sections. Duplicate forms are provided. The general procedure advised is that the children should read the first five chapters of the book, and should then take Form I of the tests. If a child passes the test satisfactorily, he reads further; if he does not he is required to read the material again and to take the second form before he may go on.

These tests represent somewhat of an innovation as regards measurement of silent reading—an innovation of some importance. They are, further, of distinct practical value to a teacher, since they permit her to determine, with very little trouble, just what her children are doing with their supplementary reading. Without such helps it is very difficult to check on such supplementary work. Finally, the work is suggestive of possible further developments along

[1]Tests have thus far been completed for *Lolami, The Little Cliff Dweller*, a story of about one hundred pages for grades II-IV, an edition of *Robinson Crusoe* of about the same size and for the same grades, and *The Story of Ulysses*, a slightly longer book for grades IV-VI.

this general line. Is it not, perhaps, desirable that most textbooks should be provided with such series of carefully devised, standardized, and convenient series of test examinations? The suggestion is commended to publishers of school books. At least, it may be said that these tests are of distinct interest; and perhaps they may lead to further developments of much importance.

Curriculum Revision in Denver

An extensive program of curriculum revision is under way in the Denver schools. This program was launched in the spring of 1922 when about forty-five committees were appointed. Most of these committees are concerned with the revision of the courses in the various subjects and grades, but some have to do with general problems of organizing and administering curricula.

In June, 1923, the Denver Board of Education, recognizing the fundamental importance of this work, made an appropriation of $31,500 for the program during the current school year. This appropriation made possible the development of a more thorough-going procedure of revision.

Of first consideration is the work of the committees of teachers. These committees have for over a year and a half been studying the problems of the selection of subject matter and methods of teaching involved in their respective fields, and have been familiarizing themselves with the results of research and experimentation wherever reports are readily available. The completed courses of study will embody the best classroom experience and judgment available in Denver. One of the chief responsibilities of the committees in a given field is that of interesting all the teachers in that field in the work in order to obtain every contribution that any teacher can make. When these courses of study are put into operation, they will be used by teachers who have had a large part in making them and who are familiar with current issues and controversies as to curricula and methods.

The appropriation made it possible for the Denver schools to enter into an agreement with the University of Colorado whereby three-fifths of the time of Professor L. T. Hopkins of the department of education is given to this program during the current school year, and with the Colorado State Teachers College whereby one-half the time of Professor W. D. Armentrout is also given to the work. The school district is paying its pro rata of the salaries of these men and the expense incidental to their weekly trips to Denver. Both these men are fitted by reason of preparation and experience for the work of curriculum revision, and they work with the committees constantly, supplying stimulation, criticism, and guidance.

When the courses have assumed tentative form, specialists in a number of fields will be brought to Denver to meet with committees for periods varying in length from five to ten days for discussion, criticism, and amendment of proposed courses. This procedure insures that the best thought in the various fields will be made available to committees.

The appropriation made by the Board of Education covers not merely the cost of the services of Professors Armentrout and Hopkins and of the visiting specialists, but also provides clerical help.

In order that members of committees may to a certain extent be relieved from their regular class work to devote the best hours of the day to this work, there is an appropriation for the employment of substitute or relief teachers.

This entire curriculum revision program is under the general direction of Mr. A. L. Threlkeld, assistant superintendent of schools, who has general charge of instruction in the Denver schools.

The appointment of a supervisor of libraries for the Denver schools, to have charge of the professional library in the Administration Building, has provided a type of library service to the department of curriculum revision that is invaluable in the preparation of bibliographies and digests.

The new Administration Building, which was occupied by the administrative staff the first of August, provides conference rooms and greatly facilitates work of this kind.

The supervisory staff of the schools is largely concerned with this problem and is making a very distinct contribution, but the organization described above relieves the supervisory staff of the immediate task of directing the work, and results therefore in the least possible interference with the administration of the schools.

JESSE H. NEWLON

Superintendent of Schools, Denver, Colorado

Smoking and Scholarship in the High School

For some time a determined effort has been made by the high-school authorities of Highland Park (Michigan) to break up the habit of smoking among the high-school boys. Of late the names of all boys known to have persisted in smoking on or near the school grounds have been collected, and their cases have been dealt with one by one. This list of names was handed to the Director of the Survey Department with a request that a study be made of the standings of these boys as compared with the standings of boys in general.

First there were excluded from the list the names of all students who had, as yet, made no permanent high-school records. The standings in academic subjects of the 84 smokers remaining on the list were then compiled from the permanent record cards. As a basis of comparison with these, the names of 105 high-school boys were selected at random, except that there were excluded the names of:

1. Boys who had recently come from other schools, excluded on the ground that the data on their cards were too meager or not comparable.
2. Senior boys and certain eleventh-grade boys seated in the senior assembly-room, excluded on the ground that they were a select group and that it would be somewhat unfair to use their standings in such a comparison.

Among the 105 were probably some smokers who would not in the comparison result to the disadvantage of the group being studied, unless smokers stand higher than non-smokers in high-school classes. The 105 boys in the

general group had on record a total of 3,090 marks in academic subjects and the 84 smokers had a total of 2,816 such marks. Classified and reduced to percents the distribution of each was as follows:

	Percent of Marks of Grade Indicated					Total
	1, or Excellent	2, or Good	3, or Average	4, or Poor	5, or Failure	
General Group of Boys	9.0	21.0	36.4	27.1	6.5	100
Group of Smokers	5.1	17.4	34.4	31.2	11.9	100

The group of smokers, as indicated in the tabulation, fell below the general group in every gradation except where they ought not to do so. A smaller percent of the group secured excellent, good, and even average marks, but when it comes to poor marks and especially to failure marks, there the group of smokers relatively outnumbered the general group by a wide margin. The general group earned over 76 percent more excellent marks, than did the smokers (as 9 is 176 percent of 5.1), but the smokers' failure marks, outnumbered the failure marks of the general group by over 83 percent (as 11.9 is 183 percent of 6.5).

In the manufacturing world the elimination of a factor that seemed to be thus associated with a possible 76 percent increase in the highest grade of output and a possible decrease of 83 percent in the scrap heap would probably attract more than passing attention. For the same reason the effect of the habit of smoking upon boys of high-school age or younger deserves more intensive and extensive study than it has received to date.

By treating the same group of marks in a different way one might evaluate the excellent, good, average, poor, and failure marks at 4, 3, 2, 1, and 0, respectively. The value of the marks of the smokers' group as expressed in the tabulation would be 172.6, while the corresponding value of the general group scores would be 198.9 or more than 15 percent above those of the smokers' group. Again it is fair to ask "What would a business firm do with a possibility of increasing the value of its total output 15 percent without a cent of additional expense?" "Find and stop the leak immediately," would probably be the order.

If the standings of the smokers were to be compared with the standings of the high-school group as a whole, including all girls as well as all boys, as shown in the semiannual reports of the principal, the result would be even more unfavorable to the smokers' group.

These facts, while not conclusive, would seem to intimate rather strongly that smoking among high-school boys is positively correlated with low scholarship marks. Whether as a cause, an effect, or just a natural accompaniment may be considered a matter of judgment or of future study. So far as this study has significance, it would seem to confirm Principal William Prakken's contention that high-school scholarship and smoking seem to be headed in opposite directions.

H. C. Daley

Highland Park, Michigan

Educational Research Association

(HARRY A. GREENE, *Secretary and Editor*)

Report of the Annual Meeting Chicago, 1924

The program of the Educational Research Association which appeared in this Department in the January number was followed with consistency. The fact that the Tiger Room of the Sherman Hotel was filled to capacity and standing room was at a premium is some indication of the amount of interest shown in the program presented by the Association on Tuesday afternoon, February 26.

The papers presented on Tuesday afternoon centered about the problems of the improvement of instruction in arithmetic. The opening discussion was presented by Director C. H. Judd of the University of Chicago School of Education. His paper dealt with the results of laboratory analysis of arithmetic. Dr. W. J. Osburn of the State Department of Public Instruction, Madison, Wisconsin, then presented data obtained from an analysis of the type and extent of drill material presented in the introductory books of a number of widely used texts in arithmetic. An eight-page mimeographed summary of the results of this analysis was distributed to the audience. The following general conclusions were stated in summary form at the conclusion of the paper:

1. Conditions in primary arithmetic are topsy turvy as far as the amount of practice is concerned. That which is easy usually receives more than twice as much practice as that which is difficult.
2. One hundred and eighty out of a total of 1,325 combinations do not occur at all in the books considered, while some easy combinations occur more than 300 times.
3. Combinations in subtraction which are prerequisite for short division are greatly neglected.
4. Great variability exists with reference to the amount and character of the practice offered in long division by six of our most widely used textbooks.
5. Further research is badly needed to determine the amount of practice which is needed to fix a combination in the minds of the children.

An excellent paper entitled "Transfer in the Narrow Mental Function of Adding Fractions" was next presented by Dr. F. B. Knight of the University of Iowa. In general Dr. Knight drew the following conclusions: In the relatively narrow mental function of adding fractions an almost complete transfer of ability seems to be obtained. Pupils in a series of tests did as many examples and did them as accurately when working with denominators they had never seen or practiced upon, as they did upon examples using denominators that had appeared in previous drill work.

Dr. Gary C. Myers of the Cleveland School of Education next presented very interesting data relative to persistence of error in arithmetic. Charts were shown which revealed a characteristic tendency of common errors in the fundamental processes to recur continually even after the attention of the subject has been repeatedly called to the error.

The final paper of the afternoon was presented by Mr. A. B. Wight, assistant superintendent in charge of measurements, Chicago.

The meeting on Wednesday morning was a joint meeting with the National Society of College Teachers of Education. The program printed in the last issue of this magazine was followed almost exactly. A discussion of the technique of trait analysis was presented by Dr. W. W. Charters, University of Pittsburgh. This was followed by a discussion of organized personnel research and its effect on high-school practice by W. H. Hughes, director of educational research, Pasadena. This discussion dealt with the results of character and personality ratings of students. The distinctive differences in method presented was that each rating covered from two to three weeks' time. A single characteristic was observed and rated during this period. Dr. W. S. Gray of the University of Chicago School of Education next presented an excellent discussion of the case method of studying reading deficiencies in the junior high school. This was followed by a paper dealing with the relation of rate and quality in educational measurements by Dean S. A. Courtis, Teachers College, Detroit. Dr. John K. Norton, director of the research division of the National Education Association briefly summarized the results of a general survey of the curriculum situation in the United States. Dr. Norton urged further emphasis on a sound philosophy of the aims and purposes of education. The results of his study led him to conclude that there must be a central agency for curriculum making. A paper entitled "Physical and Mental Development and Its Relation to Mental Age Norms" concluded the program. In this discussion Dr. F. N. Freeman, professor of educational psychology, University of Chicago, showed that the correlation of anatomical age and mental development as revealed by the Pearson r was not as high as other investigations have indicated it to be.

The closed meeting of the Association was held Thursday morning in the Crystal Room of the Sherman Hotel. With the exception of one, each person on the program was present. In view of the fact that it is planned to present in this Department resumés of the papers and discussions which were presented at this meeting the individual papers will not be reviewed here except in a very brief fashion.

Dr. M. J. Van Wagenen, University of Minnesota, presented a very detailed and somewhat complicated method of coordinating results of mental and achievement tests in school surveys. This was discussed by Mr. Joseph Peterson, Peabody College for Teachers, Nashville. In somewhat definite contrast with the technical treatment presented by Dr. Van Wagenen, Dr. M. R. Trabue submitted a graphic device for representing test scores for a period of years. This was discussed by Dr. W. W. Theisen, assistant superintendent of schools, Milwaukee. A unique graphic and mechanical method for obtaining the coefficient of correlation was next presented by Mr. E. R.

Wood, Kansas State Teachers College, Emporia. A lively discussion led by Drs. Otis, Herring, and Peterson followed the presentation of Mr. Wood's device. This discussion centered chiefly around the degree of reliability of the charts and mechanical devices. A paper prepared by Mr. H. S. Will, statistician, Toledo, dealing with the skewness and distribution of reading scores on the supervision of instruction was read by Mr. R. L. Morton, Ohio University. Dr. Ernest Horn of the University of Iowa next submitted criteria for planning and judging spelling tests. Dr. Horn's criticism of certain existing spelling tests and scales centered upon the social utility theory and the content of the scales. This report was discussed briefly by Mr. Morton of Ohio University. Superintendent C. W. Washburne, Winnetka, Illinois, reported on the merits of the individual plan of instruction as it is in operation in the Winnetka public schools. This was discussed by Dr. Rugg of the Lincoln School of Teachers College.

An interesting issue was raised in this program by Dr. Whipple in connection with the general discussion which followed. This issue dealt with the question of the kind and form of norms which should be prepared to accompany the National Intelligence Tests. The work of the preparation of new norms and the revision of available comparative data is now going forward and Dr. Whipple expressed himself as being extremely anxious to secure the cooperation of all interested persons in making these norms most useful. The question of the limitation of the age norm scale which forces many people to attempt to project the scale beyond its intended limits was raised by Dr. McCall. It was suggested by Dr. McCall that it would be much more satisfactory for Dr. Whipple and his associates to present an estimated projection of these norms than to have it done, as is now the case, by unauthorized and frequently untrained workers. In the course of the discussion Dr. Whipple made the comment that data now being compiled tended to show sex differences of considerable significance. It was the consensus of opinion that age norms should be established for both boys and girls as well as for the composite group. As a culmination of this free discussion of the general topic of the measurement of intelligence and the incidental development of an opposition to the use of the term "intelligence," it was moved by Dr. McCall and carried on the vote of the Association that the matter of terminology be referred to a permanent committee of five to be appointed by the president of the Association.

The annual dinner was held at the Hotel Sherman on Thursday evening. Forty-five members and guests attended the dinner and listened to the address of the president, Dr. Woody, who presented a careful review of the educational research of the past year. The review was based upon an analysis of articles appearing in eleven current educational magazines from January 1923 to January 1924. The types of research were classified according to six rather definite headings and the number of articles and number of pages dealing with each distinct type of research were catalogued. Dr. Woody's paper was a comprehensive and thoughtfully prepared manuscript. It was requested that it be recommended to the editor of the *Journal of Educational Research* for immediate publication.

A number of committees previously appointed made brief reports at this time. Drs. Woody and Rugg spoke briefly on the matter of the publication of an Association Yearbook, but the conclusion was reached that the organization is not at this time financially able to undertake such a publication. It was proposed, however, in a meeting of the Executive Committee that a systematic review of current research work be presented in the columns of the *Journal of Educational Research*. The responsibility for this work was not located.

At the last annual meeting of the Research Association (Cleveland) the question of the attitude of the Educational Research Association toward the affiliation of other regional associations with this body was presented by Dr. Dickson. A committee consisting of Drs. M. E. Haggerty, V. E. Dickson, and F. N. Freeman drew up the following amendments which at that time were approved by the Executive Committee and after some discussion passed by the Association by the necessary two-thirds majority:

"1. Regional organizations devoted to educational research may become affiliate members of this association upon recommendation of the Executive Committee and approval of the association.

2. Only organizations whose purpose and activities are in harmony with this association may become affiliate members, and in each case there shall be a reasonable assurance that such a regional organization gives promise of more than temporary existence.

3. Such affiliated societies may elect a representative to this association who shall be a member in good standing of this association.

4. The representatives of affiliated organizations together with the members of the executive committee shall constitute a council for the purpose of considering the problems arising out of affiliation."

Following upon this action the Executive Committee was authorized to act as a committee to report at this (1924) meeting of the organization relative to the nature of the relationships of this association to affiliate organizations and the policy which should be followed in the future. Acting on these instructions the Executive Committee through Dr. W. W. Theisen recommended and proposed a resolution which was to be substituted for the above quoted amendment. After considerable discussion the following amendment was approved by the Association:

"With the approval of the Executive Committee members of the Educational Research Association in any region may form a branch which may elect its own officers, hold meetings, elect associates of the branch and collect fees for local expenses."

Upon the adoption of this amendment it was moved and carrried that the Pacific Coast Educational Research Association members may form a branch or branches and are hereby approved by this body.

A motion from the floor called for the reading of a resolution of the Executive Committee providing for sectional meetings and round-table programs at future meetings of the Association. After some discussion the following resolution was approved:

"WHEREAS, the research interests of the members of the Educational Research Association are growing widely diversified, the Executive Committee submits the following resolution:

"*Resolved*, that at the annual meetings of the Association in addition to the open and closed programs as now carried on, the president, having ascertained the interests of members in the discussion of problems of special interest, shall organize round-table groups to provide for the desired discussion. These round tables shall be open only to members."

Following the president's address Dr. F. S. Breed emphasized the desirability of an effort being made by the Association to coordinate the research of its members to a certain extent during the coming year with the ultimate view of working out such a coordinated program in two years. The following motion was made from the floor and passed by the body:

"It is moved that the Executive Committee canvass the membership of the Association in an attempt to organize concerted research on some one problem with a view to making it available in possibly two years."

A resolution submitted by Dr. George D. Strayer and Dean Wm. S. Gray was read and adopted:

"At a recent meeting of a group of representative school men before whom the work of the Committee on Reading appointed by Commissioner Tigert was reported, it was pointed out that the assembling and interpretation of the material which should be made available for the use of teachers and supervisors would require more time and energy than the members of the committee could give while meeting their regular professional obligations. In order that this service might be rendered to the profession, it was urged that the institutions and school systems in which the committee are employed be asked to relieve them for a period of time on pay from a part or all of their regular duties in order that they may complete the work undertaken. It was recommended as well that this policy in the case of enterprises of major importance be sanctioned by professional organizations interested in such work by resolution in their next annual meeting. In support of the recommendation, several individuals reported action by governing boards under which they are employed, relieving them from their duties in order that they might render important professional service significant not only to their own group but also to the development of scientific work in education and to the efficiency of all workers in that field."

The Nominating Committee reported and the following officers were elected:

President, Dr. E. J. Ashbaugh
Vice-President, Dr. M. R. Trabue
Secretary-Treasurer, Dr. H. A. Greene

The report of the Secretary-Treasurer which follows below was approved by the Auditing Committee:

RECEIPTS

Cash on Hand (from Freeman)	$256.29	
Overrun on Cleveland Banquet	3.50	
Membership Dues (133)	532.00	
Re-instatement Fees (15)	15.00	
Total Cash	$806.79	$806.79

EXPENDITURES

Subscriptions—Public School Pub. Co.				
Old Re-instatements (12)		$ 30.00		
New Subscriptions, 1924 (118)		295.00		
			$325.00	
Executive Committee				
President's Office				
Printing	$20.30			
Postage	6.00			
		26.30		
Secretary's Office				
Printing	19.22			
Stenographic Service	2.75			
Postage	12.00			
Paper	.85			
Travel	3.76			
		38.58		
Other Expenses (MSS)				
Printing	8.75			
Miscellaneous	29.43	38.18	103.06	$428.06
Cash on Hand				378.73
Total				$806.79
Cash on Hand				$378.73
Bills Payable (Art Film Studio—Lantern Service Cleveland)				55.00
Net Assets				$323.73

No. 3

06.79

78.73
55.00

33.73